

Examining the Interaction Between Fourth Estate and Twitter: An Exploratory Case Study

Balbir S. Barn*, Ravinder Barn†, Franco Raimondi*, and Utsa Mukherjee†

*Dept. of Computer Science

Middlesex University, London, UK

Email: b.barn,f.raimondi@mdx.ac.uk

†Royal Holloway University of London

Egham, UK

Email: r.barn@rhul.ac.uk and

utsa.mukherjee.2015@live.rhul.ac.uk

Abstract—The pervasiveness of social media has resulted in increased public involvement in key discussions about social issues, as well as creating greater affordances for individual expression and collective mobilisation. In December 2012, the rape and murder of a 23-year-old Indian student in New Delhi, India, was followed by widespread condemnation and public action organised and coordinated through social media (Barn 2013). In March 2015, a controversial BBC documentary, “India’s Daughter”, about the incident was broadcast despite restrictions imposed by the Indian Government. This paper explores the interplay between mainstream media (the so-called Fourth Estate) and Twitter through a case study analysis using computational techniques to analyse 250000 tweets collated following the broadcast of the documentary. In particular, we apply the theory of postcolonialism to understand the dynamics of this interaction. Issues around implications for conducting inter-disciplinary social media research are also discussed.

Keywords—Postcolonialism; India’s Daughter; Fourth Estate; Social Network Analysis; Social Media; Twitter

I. INTRODUCTION

As the most populous democracy in the modern world, India has witnessed an increasing growth in the use of the internet in general, and social media in particular. Although accurate statistics are difficult to obtain, estimates of the micro-blogging site of 140 characters, Twitter, range from 23 million to 35 million [1]. This figure has more than doubled in the last 3 years. According to a collective called ‘India on the Internet 2014’, Twitter users in India total 35 million while 125 million people are on the social networking site (SNS), Facebook. Further, it is estimated that almost 9 out of 10 web users in India visit a social networking site. In a climate of smart phones and their applications, and SNSs such a figure is not so surprising. A core features of such social media sites is their reliance on individual users for content creation and active user involvement.

Twitter is notable in that it has rapidly become important and popular as key tool for organising and generating communication for protestors around the world [2].

In addition to the organising and communicating aspects of Twitter, researchers have also commented on how Twitter and social media in general is also being used to reconstruct and extend journalism and notions of what constitutes a Habermasian public sphere [3], [4]. That is, the realm of social life in which something approaching public opinion can be formed and where access is guaranteed to all citizens. The

outcome is that networked individuals have the capacity to use social media to enhance their role in news production and dissemination to achieve a growing independence from the Fourth Estate (mainstream media) [5].

In South Delhi, India, the 16 December 2012, rape and murder of a 23 year old physiotherapy student by six men, marked a watershed moment where some commentators asked whether this heralded an ‘Indian Spring’ [6]. The street protests across the nation in which social media was said to have played a part were described by some as the new ‘unifying force’ [7] through the formulation of a shared public opinion on social media.

Several years after the Delhi rape, a BBC documentary titled “India’s Daughter” was broadcast on 4 March 2015 in the UK and on 8th March 2015 in New York. The broadcast of the documentary directed by Leslee Udwin was controversial, in that, the Indian Government sought to have it banned and the BBC chose to bring forward the broadcast to an earlier programme slot. Early indicators of the controversial aspects of the documentary were immediately brought to the foreground. These included: the extent to which mainstream media occupied the so-called egalitarian and democratising space of social media; the postcolonial texture of the debate and the overall sense of how western media handled the case under question. Twitter naturally formed predominant backdrop to this broadcast given its initial role in mobilisation of public opinion in the original 2012 event [6], [7], [8].

Within a context where the mainstream media and social activists now largely occupy this micro-blogging space [7], research that examines the interplay between mainstream media (Fourth Estate) and social media (so called Fifth Estate) through specific case study instances can have public policy implications. Hence, the research reported in this paper makes a key analytical contribution of public reaction, through tweets, to the broadcast of the documentary on the BBC in the UK and on Youtube, Vimeo and other sources in India.

The remainder of the paper is structured as follows: In Section 2, we outline key concepts and debate around post-colonial theory given that postcolonialism was an important emergent theme. Crucially, the historical relationship between the two countries where the incident occurred and the film was shown provided a particular salience and backdrop to the analysis of this paper. In Section 3, we present the aims and details of the research methodology we have used. Section

4 presents overview results arising from the blog analysis and computational analysis of the collected tweets. Section 5 entails a discussion of some tweets followed by commentary on the validity of the results. Finally in Section 6, we present concluding remarks and outline some further research considerations.

II. RELATED WORK AND UNDERPINNING THEORY

As noted earlier, Twitter is both important and popular as key tool for organising and generating communication for protestors around the world [2]. Examples of where Twitter has played a significant role include: the Iranian protests of 2009-2010 [9], the so-called Egyptian revolution of 2011 [10] and also the various Occupy protests that took place around the world [11]. It is also clear that that the messaging technology is viewed differently depending upon temporality and context. Hence, it is seen as subversive by autocratic regimes, as well as a suitable technology for surveillance [12]. In an analysis of 104,059 tweets related to the Delhi rape incident and social protests that took place across urban India, in line with previous scholarship, Ahmed and Jaidka (2013) conclude that traditional media still plays a pivotal role in disseminating information [7]. For instance, the authors report that less than 10% of the tweets were actually from ordinary citizens / individuals. Such a finding certainly lends credence to previous observations that have questioned the egalitarian, and democratising promises of such space [13]. Questions also arise as to whether a new public sphere is being reconstructed where ordinary citizens really do have an opportunity to form public opinion. Hence, it makes sense to also provide an understanding of western media's handling of the case under question. This is particularly important to help ground the response of the postcolonial society and to also more properly explore the role of the impact of social media on public policy.

Over the last several decades, postcolonial theory has emerged as a major intellectual critical approach. The theory is generally regarded as having been founded on the contributions of key writers including Frantz Fanon, Edward Said, Gayatri Spivak, and Homi Bhabha. Primarily, postcolonial theory seeks to problematise key historical and contemporary notions, structures and processes including colonialism, race, ethnicity, culture, racism, gender, identity, inequality, and globalisation. In short, the theory seeks to 'critique and aims to transcend the structures supportive of Western colonialism and its legacies' [14]. The watershed moment for the polemics of postcolonialism was the publication of Edward Said's *Orientalism* in 1978 [15]. In *Orientalism*, Said meticulously brought out, through close textual studies, the prejudices about and biases against the non-West that informed the colonial discourse and its meaning productions. He showed how the non-West or the orient came to occupy the space of an exotic 'Other' in the canon of Western knowledge and how this 'orientalism' as a discourse was responsible in justifying the colonial and imperial projects of the Western powers. Postcolonialism, therefore, became the rallying point to challenge the presumptions of Western knowledge systems, to comprehend the epistemologies of the non-West, to create a space where, as Gayatri Chakravorty Spivak puts it "the marginal can speak and be spoken, even spoken for" [16]. Such an epistemological framework is indeed one of the key components of postcolonial theory. Others include its critique of power in

the forms of economic/cultural/economic/political/ideological domination (both historically and in the present time), its stance on the processes of otherisation, and essentialism. In the context of postcolonial theory, otherisation refers to a process by which one group uses social and psychological means to exclude or marginalise another group by focussing on differences. Whereas essentialism is understood to be the essence or "whatness" of something. In postcolonialism, essentialism implies the action of how a colonising power decides what is and isn't a particular identity. More often than not, differences and/or commonalities between groups may be overlooked to maintain a power relation.

Given this backdrop, in an examination of verbal and visual texts in United States mainstream news media reporting of the Delhi rape case, Durham argues that India / Third World is 'represented as a primitive and undisciplined space populated by savage males and subordinate women' [17]. She further asserts that in the geopolitics of sexual assault, the USA news media reinscribed social geographies of power and sex in terms of gender. Such an ethnocentric framework portrays the Third World woman as oppressed and lacking in agency, and the nation-state as incompetent and complicit in her subordination. The mediated deployment of space and place and Delhi in particular as the 'rape capital' of India serve as a key signifier of the political economy of gender and sexuality, and hence the process of ranking one society over the other.

In her analysis of over 1500 USA mainstream news articles published over a period of two months, following the December 2012 rape incident, Roychowdhury (2013) argues that through its coverage, the news reporting not only created a polarity between the new and old India within a new-liberal consumer world; but also stressed the 'notions of Western gender progressivism' as evidenced through its language including words such as 'traditional societies', 'medieval', 'rape as a weapon of power against modernity'. Here, in spite of the evidence on crimes of rape against women in the west, western space with its so-called modern cosmopolitanism is presented as safer for women. The December 2012 case is used as a platform to present a dichotomy of the modern Indian woman victim, and the backward / savage /misogynist brown man. Roychowdhury cites Spivak's 1988 writings, and argues for its ongoing appeal as witnessed in western media, that is, "white men saving brown women from brown men." [18].

We use this texturised context of postcolonialism to examine the extent to which Spivak's theoretical framework can be employed to explore the interaction between the Fourth Estate and Twitter in the context of the BBC documentary, India's Daughter. Indeed, the broadcast of this Fourth Estate film, and its discussion on Twitter provides a useful anchor to extend the postcolonial lens referred to by scholars, such as Durham and Roychowdhury.

III. AIMS AND METHODS

In this study, our primary aim was focused on the interaction between the Fourth Estate and Twitter. In doing this, we sought to apply the theory of postcolonialism to understand the dynamics of this interaction. Additional areas of interest included an identification and exploration of the debates and discussion generated as a consequence of this controversial BBC documentary. To this end, we employed a mixed-methods approach to help understand the situation namely: a series of

blogs written at the time of the broadcasting of the film in the UK, and USA in March 2015; and then the collection and analysis of tweets over a period of 4 weeks (3 March 2015 - 3 April 2015). Notably, the film was broadcast during this period to coincide with International Women’s Week. The topicality and contemporaneous nature of the study required drawing upon those social media blogs, written within a week or so of the broadcasting of the film. The blogs were read and analysed manually by two of the authors and led to the identification of the dominant themes used in the subsequent analysis. These blogs constituted not only as part of our data collection, but they were also useful in contextualising and in making sense of our Twitter data. Sixteen blogs from prominent bloggers were analysed. Most of the bloggers were female (11), and only three were male. Two of the bloggers were unknown.

A key challenge in conducting social media based research is the lack of standard approaches in appropriate methods for data collection and analysis. This concern tends to be further compounded by a limited range of integrated tools to support research methods that can enable the full range of types of analyses required. The Collaborative Online Social Media Observatory (COSMOS) is an example of a distributed digital social research platform that addresses these requirements [19]. However, at the time of this research, the tool was not readily available and furthermore did not integrate with our efforts at developing a learning set through manual analysis. Other tools such as Prometheus is a peer-to-peer service that collects social data from a number of sources and applies social inferencing techniques, but it is mostly concerned with privacy-aware social data management [20].

Given these concerns, we chose to access the Twitter data stream using the published Twitter Application Programming Interface (API) via our own bespoke software. Twitter offers a streaming API that can be filtered on keywords; in our case we employed the following list: “IndiasDaughter”, “Leslee Udwin”, “Udwin” and “banbbc”. The script was kept running to collect tweets that included the keywords from 3rd March 2015 to 3rd April 2015 following the broadcast of the documentary. Over 254,000 tweets were collected amounting to around 1GB of data. Such a volume of data requires computational approaches to analysis. Figure 1 provides an illustration of the general steps in our method.

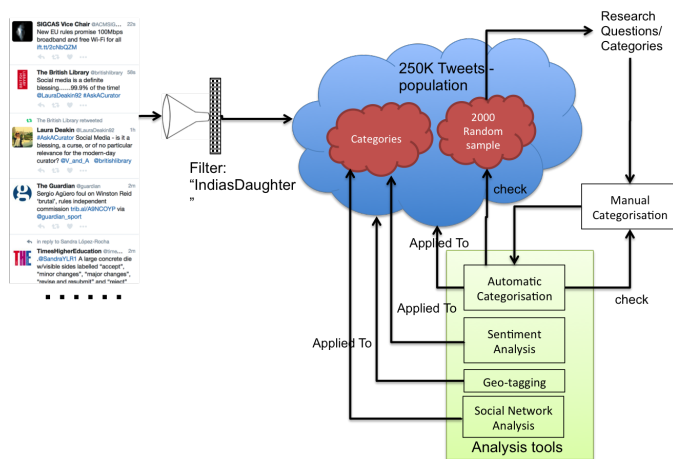


Figure 1. Method Overview.

Small scale analysis of tweets of say less than 10000 is relatively straightforwardly done by human processing. Computational approaches provide additional insights that would not necessarily be possible by manual analysis. Given the volume of the tweet set, we were interested in several types of analysis. These analytical tools included: Automatic categorisation: the use of machine learning to categorise text or other data; Sentiment analysis: the use of natural language processing and text analysis to identify attitudes of a respondent with respect to a topic; Geo-tagging: plotting the location information of tweets on a map; and Social network analysis: the use of network and graph theory to investigate social structures. We discuss the use of the analytical tools below using the diagram in Figure 1 to provide an overall context.

IV. RESULTS

As indicated earlier, our research questions have centred around several exploratory areas. The blogs, existing literature and our own research questions prompted and influenced these exploratory areas which would become coding categories in our thematic analysis of the tweets.

A. Analysis of the Blogs: Emergent themes/categories

In the analysis of blogs from March 2015, several divergences were identified. Firstly, some bloggers set out to support the ban and to justify it [21]. Others proposed taking a legal stand on the matter and argued that as the case is still subjudice, the telecast should be postponed until such time as the judgement is pronounced by the court, but in no way supporting the ban [22]. Bloggers also chose to challenge the ban, ask for it to be lifted immediately, and the telecast to take place as per the schedule [23], [24]. Others took an informed and critical stand, and commented from various perspectives such as feminism, postcolonialism or even with India’s general use of “bans”. Arguments for condemning and supporting the ban in the same breadth were presented [25], [23]. An analysis of the blogs helped generate a useful framework that could be applied to our Twitter dataset. A total of 7 prominent themes were identified. These included notions of legality of broadcasting the documentary (Legality/Ban), the postcolonial mindset of the film and the response from others (Postcolonialism), representations and discussions about the lawyers involved (Lawyers), contemporary feminist thinking in India (Feminism), the role of traditional media in discussions, in this case BBC and NDTV (mainstream media), representations and discussions about the role of Government of India (Government), and finally the role and value of punishment (Punishment).

B. Categorisation of tweets

Thematic analysis or categorisation is a powerful qualitative data analysis tool. The challenge is to deploy it for 1 GB of data. We elected to use machine learning techniques and the use of training sets. A random sample of 2000 tweets were extracted from the tweet population and classified manually against the categories listed above by two of the researchers. Where there was discrepancy, discussion was used to agree a final classification. This tweet set of 2000 tweets was used as a “Training Set” to refine / parameterise machine learning algorithms which were then used on the entire tweet population to categorise the 254K tweets. We partitioned the 2000 training

set using 3 folds. We used NLTK, a Python based toolkit for natural language processing [26]. This software comes with an open source library and toolkit for natural language processing to do stemming and tokenisation, using all the words as features. We have employed a Naive Bayesian classifier to build our model. The table in Figure 2 overleaf summarises the categorisation results. This training set had around 72% accuracy (manual versus automatic categorisation). This is consistent with other research [27]. We finally applied the model to the whole dataset of 254K tweets. Tweets related to postcolonialism amounted to 26,816, representing 10.5% of the overall total and the second largest of the analytical categories. The largest category centred around tweets about legality and banning of the broadcast. More importantly, the postcolonialism tweets amounted to 23% of the tweets that were classified against the desired classifications by the machine learning algorithm. We used the 'Other' category to denote discussion that did not fall into the categories of interest.

C. Sentiment Analysis

The use of natural language processing and text analysis to identify attitudes of a respondent with respect to a topic is a popular analytical tool used in tweet analysis. Both the original incident (through the brutality of the crime) and the subsequent controversial aspects of the televised documentary generated a wide range of emotions and efforts to assess the overall sentiment was deemed appropriate. We used the open source vaderSentiment0.5 tool [28] to conduct a sentiment analysis of the tweet data set. VaderSentiment represents sentiments on a scale from -1 to 1 representing negative sentiment at one end (-1) and positive sentiment at the other end (1). When the full set of tweets were subjected to a sentiment analysis using VaderSentiment, we found that tweets related to postcolonialism were ranked 3rd in association with negative sentiment. The overall compound sentiments for all categorisations is shown in Figure 2.

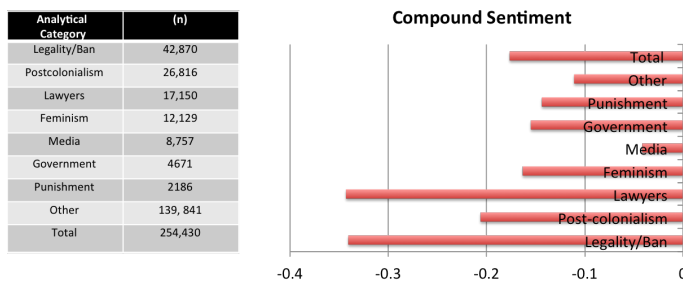


Figure 2. Automatic categorisation and overall sentiments.

D. Social Network Analysis

In this paper we are predominantly concerned with the postcolonial texture of the debate surrounding the broadcasting of the film. Hence, social network analysis on the tweet set associated with postcolonialism was conducted using two open source social network analysis tools: NodeXL [29] and Gephi [30]. The Postcolonialism tweet set (JSON file) was transformed using bespoke scripts into a form readable by NodeXL. In this data set, nodes are Twitter users and the edges represent tweets that can either be retweets or mentions. This

data set was further cleaned by merging duplicate edges and the addition of weights to reduce the edge count. NodeXL was used primarily as a means for creating the GraphML format for use in the visualisation within the Gephi toolset. Gephi is better supported on MacOS and we were able to compute and visualise various social network analysis descriptive statistics. Within Gephi, the data set comprised 2000 nodes and 13243 undirected edges. Modularity computations were performed on the network. These measured the strength of division of a network into communities. We used the modularity algorithm included in Gephi [31] and produced seven communities of interest. Each of the top four communities (in size) were centred around key Fourth Estate actors such as @BBC, @NDTV @BDUTT, @BBCIndia and @TimesOf India. Also apparent, was how these same 4th estate actors (i.e. traditional media) were also ranked highly in a range of network centrality measures such as Betweenness Centrality and Eigenvalue centrality.

Degree Centrality is a measure of a node with respect to its in-bound connections and its outbound connections. If a node/actor receives many ties they are often regarded as prominent or important. Nodes that have a high out-degree are actors that are influential.

Betweenness Centrality is a measure of node that is based on the extent to which a node falls on the geodesic paths between other pairs of nodes in the network. In social network analysis, nodes with a high value for betweenness centrality are an indication of influence on information flow in a network. Hence a node with a high value is an important conduit for information flowing between nodes in the network.

Eigenvalue Centrality considers in-bound and out-bound connections and also the node's connection to other important nodes. Hence, the measure is seen as an indicator of the power of the node.

Tiryakian et al. note that "Individuals with high betweenness centrality tend to be influential because they are well informed and can affect the flow of information in a network. As a result, they are often information gatekeepers." [32]. For example, @BBC was top-ranked for both Eigenvector (0.00929400), and Betweenness centrality (13556362.612). We also observe that the use of Eigenvalue centrality to denote power is open to debate and recent results have indicated that in Twitter, users with high eigenvector centrality need not be influential users [33]

The film was shown in both USA and the UK of which the latter has a postcolonial relationship. Analysis of geo-tagged tweets was not considered meaningful given the low numbers of geo-tagging. Other techniques such as automated analysis of Twitter handle biographies would have been a potential route for analysing how sentiments and other issues vary between the countries where the films were shown. Our future work will incorporate such approaches.

Our results for the centrality statistics are shown in table 1. These data are the non-normalised results. The top 25 results are shown. From the table it is clear that there is considerable overlap between the centrality measures of various nodes. @BBC for example is the most powerful node in the network and one with the highest betweenness measure. Several BBC based Twitter accounts feature as important conduits for information flow. Several of the nodes were Hindu nationalists,

V. DISCUSSION

TABLE I. CENTRALITY MEASURES

Vertex	Eigenvector Centrality	Vertex	Betweenness Centrality
BBC	0.00929400	BBC	13556362.612
lesleedwin	0.00635000	BBCIndia	9210704.414
YesIamSaffron	0.00344600	AlanaBowker	7989346.749
MahaveerM_	0.00344600	lesleedwin	lesleedwin
Sootradhar	0.00304900	LutyensInsider	4750346.108
ndtv	0.00301000	Sootradhar	3742481.265
TRobinsonNewEra	0.00290400	BBCWorld	2470768.572
dibang	0.00289000	gvicks	2415990.807
rishibagree	0.00261900	ndtv	2385230.780
LutyensInsider	0.00260200	vicky895	2364758.802
PMOIndia	0.00250600	dibang	2340111.494
DrJwalaG	0.00228700	shaliniscribe	2269173.050
BBCWorld	0.00228700	sunnysingh_nw3	2145012.461
thekinshu	0.00193500	M_Lekhi	1981712.913
Keisar_	0.00184800	Keisar_	1967994.765
mediacrooks	0.00169900	rishibagree	1806544.314
narendrapjoshi	0.00168700	MahaveerM_	1738846.955
mariawirth1	0.00155700	rvaidya2000	1634460.299
rvaidya2000	0.00149300	mariawirth1	1634460.299
sankrant	0.00147700	seemagoswami	1523056.837
AskAnshul	0.00143200	YesIamSaffron	1458497.885
BDUTT	0.00142300	SuchayVora	1275945.185
sunell	0.00138700	radhabharadwaj	1259731.335
htTweets	0.00134600	PMOIndia	1206559.332
rohitaarwal86	0.00129200	annieowen	1176381.602

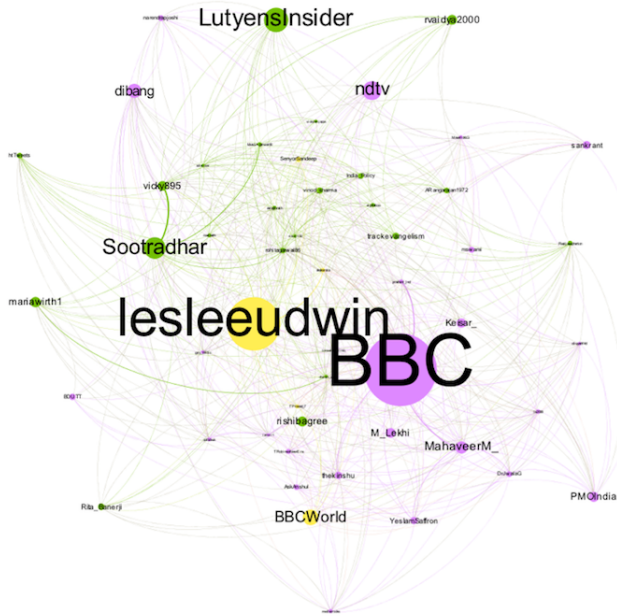


Figure 3. Power Diffusion

Others were bloggers such as @Sootradhar and @thekinshu. Presenters on Indian television programmes were represented and included @dibang, and @BDUTT for example. Most were individuals.

Figure 3 above depicts the core of the 2000 node undirected network by limiting the map to vertices (nodes) that have degree range of 25-89 edges and that also have an Eigenvalue > 0.1111762270211876 (normalised). These parameters were used primarily for presentation purposes. Additionally, the vertices are scaled by Betweenness centrality to indicate the roles that vertices are playing in brokerage and diffusion of information. The various communities to which vertices belong are also indicated by the colour.

Our results indicate that mainstream media appear to be effective in utilising the communication spaces created by social media. Thus, for example, the BBC and NDTV Twitter accounts features in both measures in Table 1. However, we can observe some impact of Fifth Estate through the centrality measures observed for some influential bloggers (@Sootradhar and @thekinshu) but noticeably not from those bloggers who were perceived to have set the agenda through their articles. The positioning of social media as a force for dissemination and a viable growing independent "Fifth Estate" as an alternative to mainstream media based journalism [34] should be questioned.

The BBC finds mention in around 50% of the tweets that were posted on the issue of the ban (categorised as Postcolonialism), but there were very few tweets from the various BBC official handles in that data set (<10). The BBC World News, with the handle @BBCWorld, on the other hand retweeted some of the tweets coming from @BBCIndia. These include the following: @BBCIndia: Letter from Director of BBC Television to Indian broadcast ministry on #IndiasDaughter. The letter that was attached to this tweet, was the one written by Danny Cohen, the Director of Television at the BBC, to the Joint secretary at the Ministry of Information and Broadcasting, Government of India. The following quote from the letter helps disambiguate the BBC’s position:

...We appreciate your concern but we feel India’s Daughter has a strong public interest in raising awareness of a global problem and the BBC is satisfied with the editorial standards of the film... The purpose of including the interview with the perpetrator was to gain an insight into the mind-set of a rapist with a view to understanding the wider problem of rape and not just in India. [35]

Further, the letter goes on to make clear whilst there is a notional respect for Indian legal jurisdiction, the BBC would bring forward the broadcast of the documentary to an earlier slot. Reasons are not given, and one possible inference is potential legal action preventing the broadcast.

...It should be noted, although the BBC is happy to take your views into consideration, we are not planning to transmit the film in any territory which lies under Indian legal jurisdiction...after lengthy and careful consideration we have decided to show India’s Daughter on BBC Four in the UK at 22:00 tonight. [35]

What is attempted in this letter has a direct bearing on our postcolonial understanding, for it redraws the battle-lines for this debate and depicts ‘rape’, the central issue of the film, as a global and not an India specific malaise. And in so doing, it attempts to steer away from any alleged bias in the representational politics despite the obvious conflation in the title of the documentary which asserts a national identity to the malaise of rape. The letter’s reference to jurisdiction appears to underplay the ubiquity and reach of social media, and seems to embolden the role of mainstream media. The Twitter discussions amplify some of the positions identified by the bloggers. Thus, Mehra notes:

There are many questions regarding why some sections of the widely-read Western media are so obsessed with reporting and analysing rapes in India while often conveniently ignoring

similarly grievous nature of the problem in their own backyards. and asks: *Can you imagine if an Indian film-maker were to make a documentary on the thousands of cases of sexual abuse and 'grooming' of young girls in Rotterham (sic) between 1997-2013, and title it "Britain's Daughter"?* [36].

The blogs and tweets from other central nodes (including bloggers and Hindu nationalists) not only accused BBC of being prejudiced in their framing of the narrative, but also came down upon the broadcaster for doing so at the expense of not giving adequate attention to gruesome incidents and issues at 'home'. For example:

@YesIamSaffron How Many Documentaries ve u Seen @BBC Reporting US Rapes or Biggest Child Exploitation Scandals in UK? #BanBBC #UHF <http://t.co/hglD7lfHOP> (297 RTs)

@thekinshu Dear @lesleedwin can you pl tell me why you afraid to make a Documentary on your own Rape Case in London.? #BanBBC

The above tweets demonstrate a strong dynamic of nationalism, and the 'other' that portrays the BBC and the documentary film maker as an oppressive force. There is a challenge to the claim made by the 'other' to be the holder of morality, and superiority. Moreover, the 'other's' stance on human rights and gender is also questioned in highlighting its own social problems. Crucially, the BBC and Leslee Udwin serve to represent the Western 'other'. In doing so, they become emblematic of the UK, the ex-colonial power and its representation of postcolonial India.

A remarkable anomaly was the hijacking of the debate by the UK anti-muslim, right-wing party leader, Tony Robinson and how Hindu nationalist Twitter users used a tweet (retweeted over 2000 times) from Tony Robinson for their own ideological stance (see @YesIamSaffron above) bringing to the foreground a curious postcolonial paradox.

@TRobinsonNewEra We are in the middle of the biggest rape and child exploitation scandal in our countries history, and the @BBC are focusing on india's rapes (2078 RTs)

It would seem that networks and how power is transferred through such networks appear to be structured by colonial and national tensions and complicated by religion, and other social divisions.

A. Threats to Validity

Internal validity issues centre around whether the micro posts (tweets) are appropriately categorised by the machine learning algorithms. Here, care was taken to use a manual process of categorisation of 2000 randomly selected tweets to help develop the training set for the machine learning software. While there is confidence that the categorisation has operated at the reference 72% accuracy [27] it is noted that a significant percentage of tweets could not be automatically categorised. Note, however, no claims on external validity (wider generalisations to different domains) are being made.

Methodological notions of validity, reliability and repeatability present a concern for much of social media research as it is challenging to be definitive that data collected from social media is actually representative of the phenomena of interest. Twitter users commenting on a particular phenomena are self-selecting. Such concerns can be partially mitigated by the use

of large scale analysis (using large data sets) but nonetheless risks such as social media being generative of the behaviours it aims to document are paramount [37]. A further concern is related to decisions around treating journalists as individual citizens or as part of the overall machinery of the Fourth Estate. We view the latter as a more representative definition.

Social media research requires inter-disciplinary thinking, but arriving at a common ground whereby a sociological theory can be adequately expressed for computational purposes is an ongoing research challenge [38].

Social network analysis, and in particular the centrality measures can offer some insight about power diffusion in networks, but as noted earlier, are open to debate. Eigenvalue centrality for example, unless correlated with inbound/outbound data may not be a good indicator of power.

Furthermore, there are limited open source seamless software tool chains that addressed the types of analyses that we utilised. There are risks in moving data between software tools. Importantly, we restrict our claims to the data that we have collected and make no generalisations for other contexts.

The ethics of using data published in social media is also of concern. The approach taken in this paper considers two key dimensions, risks of identification / disclosure to users and ethical risks around the content of the micro-blogs. The data collection (both the blogs and the tweets) are from users who would be classified as low risk users as either the user is not identifiable from a Twitter profile or is from a public, official or bot account. Ethical risks related to the content of the tweets are also limited. While the content is at times provocative and antagonistic, the classification of the users as low risk does not warrant opt in permissions before publication. Possibilities of masking identities can address confidentiality concerns, but it is important to note that Twitter Terms and Conditions state that tweets must be given in their original form and attributed to the individual who posted the tweet. Furthermore, informed consent becomes near impossible when dealing with data at large scale. The dominant use of hashtags by Twitter users also supports the notion that such users were broadcasting their thoughts specifically on a subject in a public discussion [39].

VI. CONCLUSION

Social media is becoming an important communication channel in the modern world, however its role as a democratic tool, its reach and therefore direct impact remain questionable in terms of policy formation. The recent presidential election in USA has certainly brought to the foreground, the role of social media and the need for policy discussions [40]. Our study shows that the traditional media and their components such as individual journalists continue to play a central role in these new spaces. The influence and reach of the ordinary citizen is less well pronounced so it is uncertain that the so-called Fifth Estate is really coming to the fore. Methodological concerns remain challenging. For example, common grounds whereby a sociological theory, for example, postcoloniality (as in this paper) can be adequately expressed for computational purposes remains an open question. Software tools that can support both sociological reasoning and computational analysis in a linked and coherent way has the potential to have a significant impact on addressing the issues of validity raised. Some of our future research effort is directed at developing a software tool

chain that social scientists will be able to use independently of computer scientists.

REFERENCES

- [1] I. Facts, "Topic: Internet usage in India," 2016, Retrieved: 10-07-2017. [Online]. Available: <https://www.statista.com/topics/2157/internet-usage-in-india/>
- [2] M. Tremayne, "Anatomy of protest in the digital era: A network analysis of Twitter and Occupy Wall Street," *Social Movement Studies*, vol. 13, no. 1, 2014, pp. 110–126.
- [3] J. Habermas, *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*. MIT press, 1991.
- [4] V. Belair-Gagnon, S. Mishra, and C. Agur, "Reconstructing the Indian public sphere: Newswork and social media in the Delhi gang rape case," *Journalism*, vol. 15, no. 8, 2014, pp. 1059–1075.
- [5] N. Newman, W. H. Dutton, and G. Blank, "Social media in the changing ecology of news: The fourth and fifth estates in Britain," *International Journal of Internet Science*, vol. 7, no. 1, 2012, pp. 6–22.
- [6] R. Barn, "Social media and protest - the Indian Spring?" *Huffington Post*, 2013, Retrieved: 10-07-2017. [Online]. Available: http://www.huffingtonpost.co.uk/professor-ravinder-barn/india-social-media-and-protest_b_2430194.html
- [7] S. Ahmed and K. Jaidka, "The common man: An examination of content creation and information dissemination on Twitter during the 2012 new delhi gang-rape protest," in *International Conference on Asian Digital Libraries*. Springer, 2013, pp. 117–126.
- [8] F.-M. Titzmann, "The voice of the youth, locating a new public sphere between street protest and digital discussion," *Studying Youth, Media and Gender in Post-Liberalisation India: Focus on and beyond the Delhi Gang Rape*, vol. 6, 2014, p. 79.
- [9] E. Morozov, "Iran: Downside to the "Twitter revolution"," *Dissent*, vol. 56, no. 4, 2009, pp. 10–14.
- [10] A. M. Attia, N. Aziz, B. Friedman, and M. F. Elhussainy, "Commentary: The impact of social networking tools on political change in Egypt's "Revolution 2.0"," *Electronic Commerce Research and Applications*, vol. 10, no. 4, 2011, pp. 369–374.
- [11] M. C. Stoddart and D. Tindall, "We've also become quite good friends: Environmentalists, social networks and social comparison in British Columbia, Canada," *Social Movement Studies*, vol. 9, no. 3, 2010, pp. 253–271.
- [12] E. Morozov, *The net delusion: The dark side of Internet freedom*. PublicAffairs, 2012.
- [13] F. Rebillard and A. Touboul, "Promises unfulfilled? Journalism 2.0, user participation and editorial policy on newspaper websites," *Media, Culture & Society*, vol. 32, no. 2, 2010, pp. 323–334.
- [14] J. Go, "For a postcolonial sociology," *Theory and Society*, vol. 42, no. 1, 2013, pp. 25–55.
- [15] E. W. Said, *Orientalism*. Vintage, 1979.
- [16] G. C. Spivak, *Outside in the teaching machine*. Routledge, 2012.
- [17] M. G. Durham, "Scene of the crime: News discourse of rape in India and the geopolitics of sexual assault," *Feminist Media Studies*, vol. 15, no. 2, 2015, pp. 175–191.
- [18] P. Roychowdhury, "'The Delhi gang rape': The making of international causes," *Feminist studies*, vol. 39, no. 1, 2013, pp. 282–292.
- [19] P. Burnap, O. Rana, M. Williams, W. Housley, A. Edwards, J. Morgan, L. Sloan, and J. Conejero, "COSMOS: Towards an integrated and scalable service for analysing social media on demand," *International Journal of Parallel, Emergent and Distributed Systems*, vol. 30, no. 2, 2015, pp. 80–100.
- [20] N. Kourtellis, J. Finnis, P. Anderson, J. Blackburn, C. Borcea, and A. Iamnitchi, "Prometheus: User-controlled p2p social data management for socially-aware applications," in *Proceedings of the ACM/IFIP/USENIX 11th International Conference on Middleware*. Springer-Verlag, 2010, pp. 212–231.
- [21] N. Menon, "India's Daughter and sexual violence: The issues at stake - Europe Solidaire Sans Frontieres," 2015, Retrieved: 10-07-2017. [Online]. Available: <http://www.europe-solidaire.org/spip.php?article34577>
- [22] K. Krishnan, "Rape, rape culture and the debate over India's Daughter," 2016, Retrieved: 10-07-2017. [Online]. Available: <http://scroll.in/article/711369/rape-rape-culture-and-the-debate-over-indias-daughter>
- [23] V. Narayan, "Why the documentary India's Daughter should be aired," 2015, Retrieved: 10-07-2017. [Online]. Available: <https://vinodnarayan.com/2015/03/04/should-the-documentary-indias-daughter-be-aired/>
- [24] S. Dev, "What does banning the documentary India's Daughter accomplish exactly?" 2015, Retrieved: 10-07-2017. [Online]. Available: <https://sonalidev.com/2015/03/04/what-does-banning-the-documentary-indias-daughter-accomplish-exactly/>
- [25] S. Krishnan, "India's Daughter: When men come together, but women stand divided," 2015, Retrieved: 10-07-2017. [Online]. Available: <http://www.tarshi.net/inplainspeak/indias-daughter-when-men-come-together-but-women-stand-apart/>
- [26] I. Facts, "NLTK 3.2.4 documentation," 2017, Retrieved: 10-07-2017. [Online]. Available: <http://www.nltk.org>
- [27] A. Schulz, E. L. Menca, and B. Schmidt, "A rapid-prototyping framework for extracting small-scale incident-related information in microblogs: Application of multi-label classification on tweets," *Information Systems*, vol. 57, 2016, pp. 88–110.
- [28] C. J. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Eighth International AAAI Conference on Weblogs and Social Media*, 2014.
- [29] M. A. Smith, B. Shneiderman, N. Milic-Frayling, E. Mendes Rodrigues, V. Barash, C. Dunne, T. Capone, A. Perer, and E. Gleave, "Analyzing (social media) networks with NodeXL," in *Proceedings of the fourth international conference on Communities and technologies*. ACM, 2009, pp. 255–264.
- [30] M. Bastian, S. Heymann, M. Jacomy et al., "Gephi: an open source software for exploring and manipulating networks." *ICWSM*, vol. 8, 2009, pp. 361–362.
- [31] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, 2008, p. P10008.
- [32] E. A. Tiryakian, A. Gruzd, B. Wellman, and Y. Takhteyev, "Imagining Twitter as an imagined community," *American Behavioral Scientist*, vol. 55, no. 10, 2011, pp. 1294–1318.
- [33] P. Howlader and K. Sudeep, "Degree centrality, eigenvector centrality and the relation between them in Twitter," in *Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, IEEE International Conference on. IEEE, 2016, pp. 678–682.
- [34] W. H. Dutton, "The fifth estate emerging through the network of networks," *Prometheus*, vol. 27, no. 1, 2009, pp. 1–15.
- [35] D. Cohen, "BBC - letter from Danny Cohen to Sh. Rakesh Singh regarding concerns about Storyville - India's Daughter - Media Centre," 2015, Retrieved: 10-07-2017. [Online]. Available: <http://www.bbc.co.uk/corporate2/mediacentre/statements/dc-letter-re-indias-daughter>
- [36] B. Mehra, "India's Daughter: An indian daughter's opinion," 2014, Retrieved: 10-07-2017. [Online]. Available: <http://swarajyamag.com/culture/indias-daughter-an-indian-daughters-opinion/>
- [37] A. Zubiaga, M. Liakata, R. Procter, G. W. S. Hoi, and P. Tolmie, "Analysing how people orient to and spread rumours in social media by looking at conversational threads," *PloS one*, vol. 11, no. 3, 2016, p. e0150989.
- [38] D. Murthy, "Digital ethnography an examination of the use of new technologies for social research," *Sociology*, vol. 42, no. 5, 2008, pp. 837–855.
- [39] L. Townsend and C. Wallace, "Social media research: A guide to ethics," 2015, Retrieved: 10-07-2017. [Online]. Available: http://www.gla.ac.uk/media/media_487729_en.pdf
- [40] J. Boorstin, "Facebook, Snapchat and Twitter played a bigger role than ever in the election," 2017, Retrieved: 10-07-2017. [Online]. Available: <http://www.cnn.com/2016/11/08/facebook-snapchat-and-twitter-played-a-bigger-role-than-ever-in-the-election.html>