Trolls and Social Movement Participation: An Empirical Investigation

Emergent Research Forum (ERF)

Pratyush Bharati
University of Massachusetts, Boston
Pratyush.Bharati@umb.edu

Carol Lee
University of Massachusetts, Boston
Carol.Lee002@umb.edu

Romilla Syed
University of Massachusetts, Boston
Romilla.Syed@umb.edu

Abstract
This research examines the impact of trolling behavior on individuals' participation in social movements. Troll behavior that disrupts online discourse can mislead or influence users to stray away from the main thrust of the social movement. There have been very limited information systems studies on how trolls influence social ties in specific organizational and social contexts. Current troll research is limited to the technical aspects of identifying troll characteristics. We annotated 217,582 tweets on the Stand with Jawaharlal Nehru University (JNU) movement by following a data dictionary approach and generated a word frequency for all the tweets to manually identify troll words. Researchers independently analyzed the tweets to label them based on troll characteristics. Our preliminary analysis of the troll tweet discussions elucidates the dynamics of trolls’ social influence in a social movement.

Keywords: social media, social movements, trolls, online behavior, Twitter

Introduction
Online trolling has become an important factor impacting political and social movements on social media. While the efforts of social media companies to curb online trolling cannot be discounted, the influence of trolling in society has received considerable attention in commercial media and in the public discourse. In the extant literature, few studies have focused on analyzing characteristics of troll messages (Dlala et al. 2014; Tuna et al. 2016). Recently, researchers have also begun to explore the dynamics of social movements on social media (Selander and Jarvenpaa 2016; Vaast et al. 2017). However, there is no scholarly research exploring troll behavior in the context of social movements. We contend that if social media trolling is to be contained in online movements, the characteristics of social media trolling and its effect on individual online participation has to be studied in more detail.

Our study is motivated by this knowledge gap and aims to address the research question: How does social media enabled trolling shape individuals’ online participation in a social movement? We use the case of #StandWithJNU movement that started at the Jawaharlal Nehru University (JNU), India. We conceptualize online trolling as posting unconstructive messages on social media, which are designed to provoke a reaction, to draw social media users into a fruitless argument, and to disrupt the avowed purpose of the social network (Bishop 2012). Informed by trolling and social movement literature, we analyzed the characteristics of trolling behavior and how trolling influences individual participation. Our research contributes to the growing body of research on the negative or disruptive impact of social media and the influence of social media on social movements.

Brief Literature Review
The theoretical streams that our study builds on are online trolling behavior characteristics, role of social media in social movements, and social impact of trolling.

Online Troll Behavior
Online social networks are becoming the preferred platform for trolls to post negative or offensive material due to the wide reach and potential anonymity (Cambria et al. 2010). In this paper, we have

1 Authors have contributed equally to the paper.
utilized the four main types of troll behavior criticism: (a) inflammatory, (b) abusive, (c) threatening, and (d) repetitive. Inflammatory behavior has been adapted to explain trolling in online forums (Coles and West 2016). Trolls tend to repeat destructive behavior by sending multiple messages with negative content (Seah et al. 2015). Trolls can be destructive when they are engaged in not only repetitive, but also intentional and harmful actions (Shachaf and Hara 2010).

Previous research has focused on the technical aspects of identifying trolls and their features. For example, social network analysis has been used for detecting trolls, so community managers can help stop trolling in discussion threads (Dlala et al. 2014). Certain Twitter features, such as the ratio requested to accepted friends, has been employed to categorize online users (Tuna et al. 2016). Online posts that are vulnerable for attracting trolls have properties with “a range of antisocial online behaviors that aim at disrupting the normal operation of online social networks” (Tsantarliotis et al. 2016). These techniques are useful to understand the troll characteristics adapted in this study to investigate their impact on social movements.

Social Media and Social Movements
Past literature has identified many negative factors that both deterred the growth of social movements and increased activity in protesting events, but there are no scholarly studies on the impact of trolling in the context of social media and social movements. Actual intention to protest has been studied through collective identity forms over time as participants come to share beliefs (Tremayne 2014). McAdam suggests that “scale shift”, or the process by which a small or local action becomes a major social movement can influence the organization of social media protests (McAdam and Sewell Jr 2001). These late participants tend to follow the tactics of the earlier participants and help grow the social movement (Tarrow and Tilly 2007). The identified steps of social movement participation are: become part of the potential mobilization, become a target of mobilization attempts, become motivated to participate, and overcome barriers to participation. In this study, we examine the negative factors, such as repression, which can affect social movements (Klandermans and Oegema 1987).

For example, a study on the Gulf of Mexico oil spill uses social media affordances to explain collective and organized engagement in microblogging. An analysis of the Twitter tweets suggests that there are interdependence roles, such as supporters, advocates and amplifiers, which contribute differently to the engagement. Future research can study how roles that are not dependent may affect the collective affordances (Vaast et al. 2017).

Societal Impact of Trolling
The boundary and flexibility characteristics from John Dewey’s theory of inquiry has been used to study the effects of troll attacks on a political-focused commenting platform by Gawker Media (Forestal 2017). The research found that trolls may undermine the democratic discourse and engagement in social media that differs from physical conversations. However, the specific mechanisms on the differences are yet to be researched. This study uses the social identity model of deindividuation effects (SIDE model) to understand user identity and the role in social networks, but does not focus on the impact of the deviant behavior. Due to the lack of conceptual clarity on the influence of trolls on individual social media participation, we adapt the troll behavior characteristics to explore the constraining forces of trolls in the context of social movements.

We understand the impact of trolls on social movements as three constraining forces: (a) intimidation, (b) polarization, and (c) constrains discussion. Intimidation are visible in the angry responses to trolls and can influence the discussion with negative sentiment (Hardaker 2010). The outcome of trolls to constrain online discussion is seen in provoked reactions, fruitless arguments, and disruptions of the original discourse (Bishop 2012). Troll behavior that disrupts online discourse in social movements can mislead or influence users to stray away from the main subjects of the social movement. As Dlala et al. (2014) explains, trolls intend to provoke controversy and make users deviate from the main topic of the discussion. Our study aims to understand how the behavior of trolls can intimidate, polarize and constrain discussion and, in-turn, impact participation in social movements.

Research Methodology
We adopt the case study research methodology, which is suitable to explore new topical areas and provide an in-depth understanding of the phenomenon (Eisenhardt 1989). The case study approach allows us to dive deep into the phenomenon leading to theory development and meaningful practical insights.
(Edmondson and McManus 2007; Eisenhardt 1989). We engaged in a single case design to obtain fine-grained data (Yin 1984).

**Research Context**

Stand with JNU campaign was started by the students and faculty of Jawaharlal Nehru University (JNU) in New Delhi, India. The university came into the spotlight when news media reported that an event on February 9, 2016 was held on the campus against the capital punishment of a convict and in solidarity with the struggle of Kashmiri people. Although the University had reportedly withdrawn the permission to hold the event, some student groups showed up and disrupted the event. The police arrested the JNU student union’s president, Kanhaiya Kumar, on charges of sedition along with a few other students. These arrests drew heavy criticism from media, politicians, students, scholars, global activists and civil rights organizations, and lead to a national movement for the release of students. Students organized protests, marches, talks and a hunger strike as part of this movement where they received solidarity from India and many parts of the world.

Soon after the arrests and fears from students to leave the campus, they decided to conduct their “struggle” on social media as well. The student union’s vice president voiced: “We must use our Facebook and Twitter profiles as weapons against the demonization of JNU students and spread correct information. The false narrative in the media’s voyeuristic coverage of incidents here needs to be countered with facts and sense.” While the student activists acknowledged the need to use social media to communicate their side of the story, the volunteers were also battling social media trolls. As an online campaign manager, noted: “Our aim was initially to ensure that our side of the story and facts don’t get lost in the din of Twitter... We also hope to eventually reclaim Twitter from the trolls.” The movement’s participants and activists had an active presence on various social media sites and posted regular updates.

**Data Collection and Analysis**

This study explores how troll messages impact the social movement, and, in turn, individuals’ online participation. Hence, to capture the contextual complexity (Benbasat et al. 1987), we collected data related to Stand with JNU movement using Twitter’s public API from March 9, 2016, to May 30, 2016 with relevant Twitter handles such as #StandWithJNU, @standwithjnu, and #KanhaiyaKumar. The initial sample size was 217,582 and was reduced to 153,399 after removing duplicates. We also collected the tweet author information for the 45,760 users. To indicate whether a message has troll characteristics, we annotated tweets by following a dictionary approach and generated a word frequency for all the tweets. Next, we manually identified troll words that resulted in 39,501 troll tweets. Next, each author independently analyzed the tweets to label whether the tweets have troll characteristics. As a result, the final number of troll tweets was reduced to 18,888 and the inter-coder reliability was high. Our research is exploratory in nature, thus, we adopted an inductive and iterative approach to data analysis (Klein and Myers 1999; Walsham 1995). The current literature on online trolling and social movement served as a “sensitizing device to view the world in a certain way” (Klein and Myers 1999). The interpretive approach not only allowed us to conduct the data analysis within the purview of prior theories (Leong et al. 2015; Tim et al. 2017), but also allowed us to iterate between data and theory to identify new themes and constructs (Charmaz 2006; Glaser and Strauss 1967). To explore the impact of troll messages on users’ intention to participate in social movement, we transformed the data to identify the tweets posted in response to original troll tweets. Hence, for each troll tweet, we may have multiple response tweets. The data analysis continued until we reached theoretical saturation (Eisenhardt 1989). For further analysis, we will adopt a narrative analysis of Twitter postings to provide a thick description of events (Langley 1999). This will allow us to identify the sequence of events and delineate the phases of the process model from triggering mechanisms (i.e. trolling in our case) (Langley 1999).

**Preliminary Findings**

Our data analysis suggested several troll characteristics related to the three theoretical constructs: intimidation, polarization, and constrains discussion. Table 1 presents a summary of the constructs along with exemplar tweets. In the analysis of troll tweet discussions, we identified the impact of troll messages on user’s intention to participate in the movement. Our preliminary findings show the effect of troll messages on the intention to participate in a social movement.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Definition</th>
<th>Troll Characteristics</th>
<th>Exemplar Tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intimidation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polarization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constrains discussion</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Trolls and Social Movement Participation

<table>
<thead>
<tr>
<th>Intimidation (Hardaker 2010; Shachaf and Hara 2010)</th>
<th>Troll’s provocation elicits the desired angry response, and can be destructive when they are engaged repetitive, intentional and harmful actions</th>
<th>The posts indicate inflammatory, abusive, and threatening criticisms</th>
<th>“those who placed #kanhaiyakumar on a pedestal must feel fairly foolish after revelations about his delinquent behavior”; “do leftist dogs have anything left to bark on their anti-nationalism anymore when its proved now they are guilty”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polarization (Lee et al. 2014)</td>
<td>Troll’s continue posting of intentional disruption and abusive messages can escalate online discourse and increase polarization</td>
<td>These posts aim to polarize the discussion on the lines of gender, religion, and ideology.</td>
<td>“A woman has been walking on the streets of Jaipur [city] in her lingerie and calling it ‘azaadi’ [independence] a la lady”; “no brahmin will even touch this buffoon”; “@shehla_rashid kindly disclose your list of terrorists with name, religion, sex in Kashmir”</td>
</tr>
<tr>
<td>Constrains Discussion (Bishop 2012)</td>
<td>Troll’s desire to provoke a reaction, to draw targets (and others) into fruitless argument, and to disrupt the avowed purpose of the group gathering</td>
<td>The discussion is constrained through expression of disgust and mockery</td>
<td>“i just wonder how many years one student takes to finish his course #dramaqueen #kanhaiyakumar”; “@shehla_rashid taxpayer 4m when weeding over taxpayers money for yrs in jnu,shameless scums”; “and pseudo intellectuals of this country compare #kanhaiyakumar with our shaahed-e-azam #bhagatsingh #shame”; “we live in a country where rich people like priyanka gandhi is unable to play rent &amp; poor people like #kanhaiyakumar travel”</td>
</tr>
<tr>
<td>Intention to Participate (Klandermans and Oegema 1987; Tarrow and Tilly 2007)</td>
<td>Participant’s intention to become part of the potential digital or real protest, become motivated to participate, and overcome barriers to participation</td>
<td>The posts suggest the users interest in participating in digital or real protest or simply organizing the protest</td>
<td>“1000s join the people’s march to #savedemocracy from mandi house to jantar mantar in new delhi. #standwithjnu”; “join jnusu-jnuta protest demo at 4 p.m. at ad block, jnu. burning copies of the hlec report”; “day 3-indefinite hunger strike in jnu against political witch-hunt of students by rss-administration.”; “the socially backward dont have to bear the brunt of corporate loot #marchforazaadi #standwithjnu @cpimspeak”</td>
</tr>
</tbody>
</table>

Table 1. Troll Characteristics and Exemplar Tweets

Our preliminary findings provide theoretical clarity regarding the constraining forces of trolls pertaining to social media based negative and disruptive behaviors on social movements. Using an inductive and iterative approach to data analysis, we found that troll messages that exhibit certain troll characteristics can provide clarity about the effects of users’ intention to participate in the social movement. These constraining forces can be used to understand the impact of trolls in online discussions. Our study is likely to contribute to social media and social movement literature by examining the characteristics and the impact of troll messages on social movements. On a practical level, the findings will help social movement organizers better manage social media trolls. The findings could also benefit policy makers in managing provocative and abusive trolls on social media.

REFERENCES