

Social Ties That Mobilize: Predicting Sustained Online Protest Participation Through Network Metrics

Kokil Jaidka[1(✉)] and Saifuddin Ahmed[2]

[1] Department of Communications and New Media, National University of Singapore, Singapore, Singapore
jaidka@nus.edu.sg

[2] Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore, Singapore

Abstract. Social media has become instrumental in mobilizing around political issues and social movements. While many studies have examined the linguistic trends in protester mobilization, there is limited analysis of the network dynamics that support sustained protest participation. This paper presents a cross-national multi-method analysis focusing on social network metrics and their relationship to protester attrition in two datasets: the #Nirbhaya protests in India and #BlackLivesMatter in the USA. Our findings reveal that network position metrics – particularly degree centrality, eigenvector centrality, and PageRank – are strong predictors of sustained participation, often outperforming emotion-based predictors. Users with higher degrees of connectivity and centrality within the protest network are significantly more likely to remain engaged throughout the movement. These results suggest that the formation of social ties within protest networks serves as a crucial mechanism for maintaining activist commitment beyond initial emotional reactions. We conclude with theoretical implications for understanding collective identity formation in digital activism, implications for protests in an AI-mediated, high-misinformation and high-polarization social media environment.

Keywords: social media · protest networks · #blacklivesmatter · India · USA · Twitter · social network analysis · collective identity · protester attrition

1 Introduction

Social media has fundamentally transformed how social movements are mobilized and sustained. The open, pervasive nature of platforms like Twitter provides unprecedented opportunities for civic activism by enabling rapid information dissemination and social coordination [7,8]. While numerous studies have investigated the role of social media in contributing to offline protests [11–13,19],

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
O. Gervasi et al. (Eds.): ICCSA 2025, LNCS 15650, pp. 53–67, 2025.
https://doi.org/10.1007/978-3-031-96962-1_4

most have focused on aggregate trends in emotions and semantics rather than the underlying network structures that may determine protest longevity.

The central limitation of previous approaches is their tendency toward ecological fallacy – where aggregate emotional trends across protesters may not reflect individuals' social journeys throughout online social movements. Furthermore, these approaches often overlook a critical dimension: the evolving patterns of social connections between protesters and how these network positions influence sustained participation.

This study addresses this gap by asking: *How well do network metrics derived from early social media interactions predict individuals' likelihood of sustained participation in an online protest?* We operationalize protest participation as a binary variable indicating continued tweeting about the protest, with our primary predictors being social network metrics including degree centrality, eigenvector centrality, closeness, and PageRank, along with emotional and cognitive language features for comparison.

Our research offers three key contributions: First, we provide theoretical insights into how network position influences collective identity formation in social media movements across different cultural contexts. Second, we present a methodological contribution through a longitudinal design that uses network metrics to study sustained protest participation. Finally, we reconsider assumptions about emotional drivers of protest by demonstrating how network positions can sometimes be more predictive than emotional expressions.

2 Related Work

Social media platforms facilitate protest mobilization by promoting participatory behavior [4,46], instigating public discussion [14], disseminating information [45], and creating new connections [42]. Many modern protests originate online before moving offline [18,20], often with significant political implications [27].

Studies of network dynamics in social movements have yielded mixed findings. Conover et al. [10] found that highly interconnected users with pre-existing interests initiated the Occupy Wall Street movement online but tended to lose interest over time, suggesting pre-existing social bonds may not sustain movements. Conversely, Valenzuela et al. [46] demonstrated that online social ties are crucial for understanding both online and offline protest participation over time. These findings align with broader theories of connective action, where networked affordances of social media platforms allow individuals to mobilize without centralized leadership [4]. Social media enables networked framing processes in which activists, journalists, and audiences collaboratively construct protest narratives, often through ambient, participatory conversations [30,31,39].

Despite this evidence that network structure matters, most research has prioritized content analysis over network analysis. Studies have explored how social media messaging – particularly emotional and linguistic cues – mobilizes participation in online protests. Ahmed, Jaidka, and Cho [1] conducted a text-based emotion analysis of protest discourse in New Delhi, finding that offline emotions corresponded with their online build-up. However, their operationalization

of collective identity relied solely on pronoun usage, missing the opportunity to leverage social network interactions as indicators of social ties and collective identity. Similarly, other studies have examined emotional trends in various protests [33,36,38] without addressing individual-level network positions.

Analyses of the #BlackLivesMatter movement have explored demographic features, geographic distribution, and relationships between online and offline protests. De Choudhary et al. [11] found that social media protest volume predicted future street demonstrations, with changing engagement and topics over time. However, these results were reported at an aggregate level and did not illuminate factors determining individuals' continued involvement. Freelon et al. [16] found that "commitment" (regular posting) was the strongest predictor of elite attention to protest issues, demonstrating the importance of sustained activism, yet the network determinants of individual commitment remained unexplored.

The current study focuses on individual users who posted on Twitter before, during, and after two social media protests: the #Nirbhaya protests following a gangrape incident in New Delhi (December 2012) and the #BlackLivesMatter protests following Michael Brown's death in Ferguson, USA (August 2014). Given social media's vital role in mobilizing citizens during political protests, understanding which aspects of social network position best predict continued participation offers critical insights for both movement organizers and researchers.

2.1 Network Position and Sustained Protest Participation

Collective identity enables people to act as group members rather than individuals [47], transforming individual action into a community consciousness that drives collective action [21]. Teo and Loosemore (2017) argue that "a protest group might be made up of multiple and shifting subgroups and communication networks all working together in different variations towards a loosely defined end goal." These networks play a crucial but understudied role in sustained participation.

While emotional reactions of protesters have been extensively studied as drivers of participation [9,23,24], the structural characteristics of protesters' positions within social networks have received comparatively less attention. Limited prior work suggests that network centrality metrics may be powerful predictors of sustained engagement. Individuals with higher PageRank or eigenvector centrality would occupy more central positions in information dissemination networks, potentially increasing their commitment to the movement [18].

Drawing on these insights, we hypothesize that social media users who develop stronger network ties with other protesters would be more likely to maintain their participation. These network connections should serve as structural manifestations of collective identity, binding individuals to the movement beyond initial emotional reactions. We expect that protesters with higher centrality metrics will sustain their commitment to online protests longer than those with peripheral network positions, regardless of emotional expression patterns.

3 Methodological Framework

The following section describes our approach to extracting network and comparative features from social media posts, followed by our predictive modeling methodology and dataset descriptions.

3.1 Feature Extraction

For each user, we calculated network and comparison features for three time periods: (a) Before the protest, (b) During the protest, and (c) After the protest.

Demographic Features. We used the language-based predictive model provided by [40] to predict the age and gender of a 30% sample of individuals in either dataset. The mean and median age in both datasets was 28 years. 55.8% of individuals in the Nirbhaya dataset and 77.2% in the #BLM dataset were male.

Textual Features for Comparison. We extracted emotional and cognitive measures based on previous literature identifying their role in protest participation [9,24]. We included these as comparison features to assess the relative importance of network position versus emotional expression. Furthermore, including textual features aligns with findings from Mendelsohn et al. [29], who show that social media movements often use diagnostic, prognostic, and motivational framing to structure collective sense-making. Their large-scale analysis of movements in the U.S. context illustrates how framing tasks can evolve in response to interactional and organizational roles, reinforcing collective identity through message strategies such as calls to action and expressions of solidarity. Accordingly, we considered the following textual features:

N-gram Extraction: We used a bag-of-words representation to reduce each user's posting history to a normalized frequency distribution of unigrams, bigrams, and trigrams, retaining only those used by at least 1% of users:

$$freq_{rel}(user, ng) = \frac{freq(user, ng)}{\sum_{ng' \epsilon ngs} freq_{abs}(user, ng')} \quad (1)$$

Using standardized psycholinguistic lexica – Linguistic Inquiry and Word Count (LIWC) [37] and the EmoLex emotion lexicon [32] – we extracted emotional and cognitive language features at the post level, then aggregated these into period measurements. Following [1], we removed topical words denoting negative emotion to avoid biasing our analysis. Categories included:

- **Emotional appraisal**: Words denoting positive emotion (joy, anticipation) and negative emotion (anger, anxiety, disgust, fear, sadness).

- **Cognitive appraisal**: Words denoting insight (*think, know*), causation (*because, effect*), discrepancy (*should, would*), tentative (*maybe, perhaps*), certainty (*always, never*), differentiation (*hasn't, but, else*) and perception (*look, heard, feeling*).
- **Time orientation**: Words reflecting past focus (*ago, did*), present focus (*today, now*) and future focus (*maybe, will*).
- **Identity words**: Words depicting self (*I, my*) or collective identity (*we, our*).

Social Network Metrics. Our primary focus was on network metrics derived from user interactions. For each individual, we calculated:

- **Degree centrality**: The total number of unique users mentioned by or mentioning the individual, representing their direct connectivity within the protest network.
- **Eigenvector centrality**: A measure of influence that considers not just the number of connections but also the centrality of those connections. It is proportional to the sum of the centralities of users with which an individual is connected:

$$C_e(x_i) = \frac{1}{\lambda} \sum_j A_{ij} x_j \tag{2}$$

  where A is the adjacency matrix of the mention network, λ is a constant, and x_j is the centrality of user j.
- **Closeness centrality**: A measure of how near an individual is to all other individuals in the network:

$$C_c(x_i) = \frac{1}{\sum_j d(x_i, x_j)} \tag{3}$$

  where $d(x_i, x_j)$ is the shortest path distance between users i and j.
- **PageRank**: A recursive centrality measure capturing how often a user would be encountered in a random walk through the mention network:

$$C_p(x_i) = \alpha \sum_j A_{j,i} \frac{C_p(x_j)}{d_j^{out}} + \beta \tag{4}$$

  where A is the adjacency matrix, d_j^{out} is the out-degree of user j, and α and β are constants.
- **Mention network density**: The ratio of actual connections to potential connections among a user's immediate network neighborhood.
- **Structural holes**: The absence of connections between a user's neighbors, calculated using Burt's constraint measure.

We intentionally did not use friend and follower counts, which reflect pre-existing peer networks but may be unrelated to protest participation, as suggested by prior work showing these pre-existing ties may not predict continued participation [25], and by findings that argue that frame construction on social media is deeply shaped by message-level interactions, such as replies and quotes, which are often more diagnostic in nature and engage audiences beyond the immediate follower network [29].

3.2 Predictive Modeling

We considered an individual's daily tweet volume about the protest as an indicator of participation. Setting a minimal activity requirement ensured our sample population was comparable in posting frequency until the actual protest event. We classified an individual posting less than one protest-related post per day during or after the protest event as a dropout.

We approached prediction as a binary classification task, using support vector machines (SVMs) with a Gaussian radial basis function kernel to identify predictors of sustained participation:

$$K(x_i, x_j) = exp(-\gamma||x_i - x_j||^2), \gamma > 0 \quad (5)$$

where x_i and x_j are feature vectors and γ is the free parameter.

We trained separate models using: (1) prior participation features alone, (2) emotional language features, (3) cognitive language features, (4) network position features, and (5) combined features. This approach allowed us to compare the predictive power of network metrics against traditional content-based features.

4 Data

We analyzed two datasets: the Nirbhaya dataset from India and the #BlackLivesMatter dataset from the USA. Figure 1 shows how the proportion of protesters changed over time in both datasets, with a sharper drop in the Nirbhaya protest compared to more sustained participation in the #BLM movement.

4.1 The Nirbhaya Dataset

The Nirbhaya dataset [1] comprises 71,306 tweets posted during three protest stages: before (Dec 17–21, 2012), during (Dec 22–23, 2012), and after (Dec 24–27, 2012). Tweets were collected matching keywords related to the Delhi rape case. After filtering non-English tweets, 65,613 tweets remained.

We identified users who posted daily about the incident before the protest and at least five times during and after, excluding news accounts and aggregators. From 556 users meeting the activity criterion, 361 disassociated from the protest (dropouts), contributing 25,818 pre-protest messages. The remaining 169 protesters posted consistently throughout, with 119 continuing after the protest. These committed protesters contributed 26,820 pre-protest, 30,072 during-protest, and 31,493 post-protest messages.

4.2 The #BlackLivesMatter Dataset

The #BlackLivesMatter dataset [11] contains tweets from the first wave of protests (Aug 8–27, 2014), with over 10 million posts from 2.1 million users. After filtering for English tweets, we identified users posting daily before (Aug

8–13), during (Aug 14–20), and after (Aug 21–27) the protest, excluding news accounts and focusing on original content rather than retweets.

From 2,938 users posting daily before the protest, 1,382 disassociated at some point (dropouts), contributing 217,562 messages. The remaining 1,556 protesters continued posting throughout, contributing 333,717 messages during the observation period.

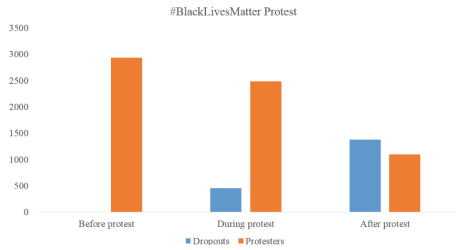


Fig. 1. Characteristics of drop-outs and protesters in the #BlackLivesMatter dataset. Blue represents dropouts; Orange represents protesters. (Color figure online)

5 Results

5.1 Predictive Performance of Network Metrics

We report predictive performance in terms of area under the ROC curve (AUC) for different feature sets (Fig. 2). Contrary to expectations, prior tweet frequency was among the weakest predictors, especially for the Nirbhaya protest (AUC (#BLM) = 0.70; AUC (Nirbhaya) = 0.59).

The most significant finding was the strong performance of network position features, which achieved AUC = 0.78 for #BLM and AUC = 0.74 for Nirbhaya. This was comparable to and in some cases exceeded the performance of emotional features (AUC (#BLM) = 0.75; AUC (Nirbhaya) = 0.83). The difference in performance between cognitive features was substantial between datasets (AUC (#BLM) = 0.64; AUC (Nirbhaya) = 0.81).

When examining specific network metrics, degree centrality and PageRank were consistently the strongest predictors across both datasets. Combined models using all features achieved the highest performance (AUC (#BLM) = 0.84; AUC (Nirbhaya) = 0.87), suggesting that network position and psychological factors provide complementary signals about protest commitment.

5.2 Network Metrics and Behavioral Insights

Table 1 identifies the most significantly correlated features from each category, based on Bonferroni-corrected Pearson correlations. Among demographic

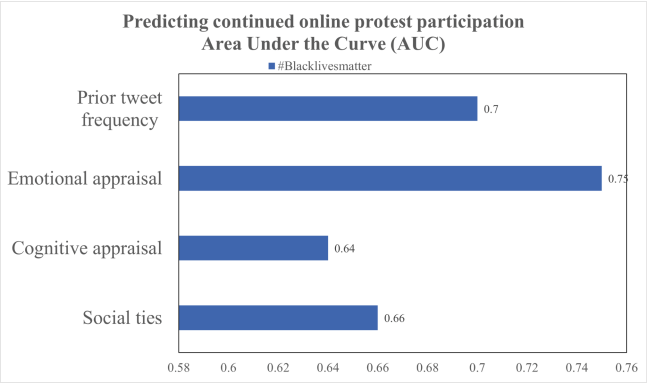


Fig. 2. Predictive performance of SVM classifiers trained on the #Nirbhaya and the #BlackLivesMatter datasets. Network position features show strong predictive performance, comparable to emotional features.

features, age negatively correlated with protest dropout in both datasets ($r(\text{Nirbhaya}) = 0.13$, $p < 0.05$; $r(\#\text{BLM}) = 0.20$, $p < 0.05$).

The most revealing findings came from the network metrics analysis. Degree centrality was the strongest network predictor of sustained participation ($r(\text{Nirbhaya}) = 0.13$, $p < 0.05$; $r(\#\text{BLM}) = 0.11$, $p < 0.05$), indicating that users who interact most with other protest participants were most likely to continue their involvement. PageRank showed similarly strong correlations ($r(\text{Nirbhaya}) = 0.11$, $p < 0.05$; $r(\#\text{BLM}) = 0.13$, $p < 0.05$), suggesting that users occupying influential positions within the protest network had higher commitment levels.

Interestingly, closeness centrality predicted dropout in the #BLM protest ($r(\#\text{BLM}) = -0.15$, $p < 0.05$). This counterintuitive finding may identify users who were "broadcasters" lying on short paths to many other users but potentially within isolated groups. These users may have een influential within their small communities but lacked the broader network connections necessary for sustained involvement.

The network density around users also showed differences between protests. In the Nirbhaya case, protesters with denser local networks were more likely to continue participation ($r(\text{Nirbhaya}) = 0.09$, $p<0.05$), while in the #BLM case, network density was not significantly correlated with continued participation.

For comparison, among emotional features, anger was a significant predictor of sustained participation ($r(\text{Nirbhaya}) = 0.18$, $p < 0.05$; $r(\#\text{BLM}) = 0.17$, $p < 0.05$). Among cognitive features, insight was consistently significant ($r(\text{Nirbhaya}) = 0.10$, $p < 0.05$; $r(\#\text{BLM}) = 0.14$, $p < 0.05$). Self-references were negatively associated with participation in the Nirbhaya dataset but positively in the #BLM dataset ($r(\text{Nirbhaya}) = -0.08$, $p < 0.05$; $r(\#\text{BLM}) = 0.11$, $p < 0.05$), suggesting cultural differences in protest discourse.

Table 1. Statistically significant features associated with dropout. The category in blue are predictive of continued participation; categories in red are predictive of dropout ($p < 0.05$).

Concept	Category	Example words	#BLM dataset	Nirbhaya dataset
Demographics	**Age**	-	0.20	0.20
Emotional appraisal	**Anger**	hate, hell, fuck, stupid, sucks, crap	0.17	0.18
Cognitive appraisal	**Insight**	think, know, understand, believe	0.14	0.1
	Compare	like, more, worse, worst	0.19	-
	Past focus	was, got, had, been, did, were	0.13	-
Identity	**1st person singular**	i, my, me, i'm, im, i've, myself	0.11	-0.08
	1st person plural	we, our, us, let's, we're, lets	0.10	
Social ties	**Closeness centrality**	-	-0.15	-
	Degree centrality	-	0.13	0.11

6 Discussion

Our findings reveal a critical yet understudied dimension of online protest dynamics: the predictive power of network position in determining sustained participation. The strong performance of network metrics-particularly degree centrality ($r(\text{Nirbhaya}) = 0.13$, $r(\#\text{BLM}) = 0.11$) and PageRank ($r(\text{Nirbhaya}) = 0.11$, $r(\#\text{BLM}) = 0.13$)-demonstrates that structural embeddedness within protest networks is often as important as emotional expression in predicting commitment to social movements. This challenges conventional approaches that prioritize aggregate content analysis over structural network dynamics.

The role of degree centrality specifically highlights how reciprocal interactions create accountability structures that reinforce continued participation. Users who establish multiple connections with other protesters effectively develop a social scaffold that sustains their involvement beyond initial emotional reactions. This finding substantiates Valenzuela et al.'s [46] argument that online social ties are crucial for sustained participation, while extending it by identifying specific network metrics that quantify these relationships. The significant predictive power of PageRank further reveals that individuals who become reference points within protest networks-those who are frequently mentioned and replied to-experience increased pressure to maintain their commitment, perhaps due to perceived leadership responsibilities.

Interestingly, we found divergent patterns for closeness centrality, which negatively predicted sustained participation in the #BlackLivesMatter protest ($r = -0.15$). This suggests that users functioning primarily as information bridges or broadcasters-those who can quickly reach many others but lack dense reciprocal connections-may be less invested in the movement itself. They may serve as amplifiers without developing the personal commitment that stems from deeper network integration. This nuance helps reconcile conflicting findings in previous literature, particularly Conover et al.'s [10] observation that interconnected users with pre-existing interests often lose interest over time. Our results suggest that it is not merely connectivity but the type of connectivity that matters-dense, reciprocal ties promote commitment more effectively than bridge positions.

The cross-cultural consistency of our findings between Indian and American contexts is particularly striking. Despite substantial differences in political systems, cultural norms, and protest traditions, the fundamental relationship between network integration and sustained participation remains remarkably similar. This suggests that the structural mechanisms of collective identity formation in digital spaces transcend cultural boundaries, even as the specific emotional and cognitive expressions may vary between contexts. The differential patterns in self-reference usage-negative association with participation in Nirbhaya ($r = -0.08$) versus positive association in #BLM ($r = 0.11$)-highlight how cultural differences manifest in protest discourse while underlying network dynamics remain consistent.

Our multi-method approach also enables us to reconcile seemingly contradictory theories of protest mobilization. While confirming the mobilizing power of anger ($r(\text{Nirbhaya}) = 0.18$, $r(\#\text{BLM}) = 0.17$) that aligns with Jasper's [24] and Castells' [9] emphasis on emotional drivers, we demonstrate that these emotions operate alongside network processes rather than in isolation. The temporal dimension of our analysis suggests a potential mechanism: initial participation may be emotionally driven, but sustained commitment depends on network integration that transforms individual emotional reactions into collective identity. This supports Bennett and Segerberg's [4] theory of connective action while providing empirical metrics to measure this connective tissue in online movements.

The strong performance of combined predictive models (AUC (#BLM) = 0.84; AUC (Nirbhaya) = 0.87) indicates that network position and psychological factors provide complementary rather than competing explanations for protest commitment. This integrated approach offers a more holistic understanding of online activism that recognizes both what protesters say and how they connect with others as crucial determinants of movement longevity.

7 Conclusion and Implications

This study advances our understanding of digital activism in three significant ways. First, it provides empirical validation for theories of collective identity formation by demonstrating how structural network positions-quantified through specific metrics-predict sustained participation. The strong performance of network features (AUC = 0.78 for #BLM; AUC = 0.74 for Nirbhaya) reveals that the architecture of social connections serves as a crucial scaffold for maintaining protest commitment beyond initial mobilization.

Second, our research offers methodological innovations by combining network analysis with computational linguistics in a longitudinal design. This approach transcends the limitations of aggregate content analysis that dominates current research, enabling prediction at the individual level and avoiding ecological fallacies. The superior performance of combined models suggests that future research should integrate structural and content analyses rather than treating them as separate approaches.

Third, our findings have practical implications for movement organizers and platform designers. For organizers, fostering dense interaction networks among

participants may be as crucial for sustaining momentum as crafting emotionally resonant messages. Our results suggest that creating opportunities for reciprocal engagement-encouraging replies, mentions, and multi-directional conversations rather than one-way broadcasting-may help transform emotional reactions into durable commitment. For platform designers, these findings highlight how algorithmic amplification that prioritizes bridge nodes or broadcasters over densely connected communities may inadvertently undermine the structural foundations of sustained activism.

7.1 Implications for an AI-Mediated, High-Misinformation Environment

Our findings take on heightened significance in today's evolving digital landscape, characterized by increasing bot activity, AI interventions, misinformation campaigns, and affective polarization. The predictive power of network metrics provides both warning signals and strategic opportunities for authentic social movements operating in such contested spaces.

First, our findings on network integration highlight a potential vulnerability in online movements. As Varol et al. [48], Ng and Carley [34], and Luceri et al. [28] have demonstrated, coordinated networks of bots and inauthentic accounts can simulate the structural patterns of genuine activism, potentially triggering the same mechanisms of perceived collective identity that sustain authentic participation. This suggests that movement organizers must develop more sophisticated methods for verifying the authenticity of new participants, particularly those who rapidly achieve high centrality metrics [15].

Conversely, the same network insights offer a potential defense against manipulation. Our findings suggest that genuine online movements develop characteristic patterns of reciprocal engagement that may be difficult for automated systems to authentically replicate. As Broniatowski et al. [6] observed in their analysis of vaccine debates, bot accounts typically excel at information broadcasting but struggle to maintain the complex, emotionally nuanced interactions that our study identifies as crucial for sustained participation. This creates an opportunity for movement organizers to foster interaction patterns that are resistant to artificial amplification-what Starbird et al. [41] term "participatory resilience."

The rising tide of misinformation introduces additional complexities. Tucker et al. [43] and Bradshaw and Howard [5] have documented how state and non-state actors strategically target social movements with misleading content designed to foster division or redirect activism toward unproductive targets. Our network-based approach offers a potential counterweight to such tactics. By focusing on building dense, reciprocal networks with strong collective identity markers, movements may develop greater resistance to divisive misinformation, as Benkler et al. [3] observed in networked communities that successfully maintained epistemic integrity despite exposure to contradictory information flows.

The problem of affective polarization [22] presents perhaps the most significant challenge to online movements. As Bail et al. [2] demonstrated, increased exposure to opposing political views on social media can paradoxically increase

rather than decrease polarization. Our finding that sustained participation depends on network integration suggests that movements seeking broad-based support must navigate a delicate balance-creating sufficient network density to sustain commitment while remaining permeable enough to incorporate diverse perspectives. This may require what Tufekci [44] calls "tactical freeze," deliberately maintaining ambiguity in some movement objectives to allow coalition-building across potential divides.

In an environment increasingly populated by AI-generated content [49], the interaction patterns we identify as predictive of sustained participation may become crucial markers of authentic human engagement. As Lazer et al. [26] argue, addressing the challenges of misinformation and manipulation requires focusing not just on content but on the structural features of information exchange. Our findings contribute to this effort by identifying specific network signatures of genuine, sustained activist commitment that may help distinguish authentic movements from orchestrated campaigns.

7.2 Future Directions

This work opens several promising avenues for future research. Longitudinal studies examining how network metrics evolve throughout the lifecycle of social movements could provide deeper insights into the dynamics of collective identity formation. Cross-platform analyses could explore whether these network effects persist across different social media environments with varying affordances. Additionally, examining how online network positions translate to offline participation could address questions about the relationship between digital and physical activism.

Future studies should also develop more sophisticated methods for measuring collective identity formation through combined network and linguistic features, building on our finding that these approaches provide complementary signals. Recent computational techniques for frame detection developed by Mendelsohn et al. [29] could be integrated with network metrics to provide more granular insights into how collective identities are negotiated and maintained in online spaces.

Given the growing concern about inauthentic participation in social media movements, future work should also investigate how the network signatures we identify differ between organic movements and those subject to coordinated manipulation. Techniques for detecting coordinated inauthentic behavior proposed by Nizzoli et al. [35] and Giglietto et al. [17] could be combined with our network-based approach to develop more robust movement authenticity metrics.

The limitations of our study include the focus on English-language tweets, which may underrepresent non-English speaking participants, particularly in the Indian context. Additionally, while our demographic models provide useful controls, they may have differential accuracy across cultural contexts. Future work should incorporate advances in multilingual sentiment analysis and demographic inference to address these limitations.

In conclusion, our findings reveal that the architecture of social ties in online protests creates structural foundations for sustained activism. While emotional triggers may spark initial participation, it is the integration into dense, reciprocal networks that transforms momentary outrage into enduring commitment. As online spaces become increasingly contested through automated manipulation and strategic misinformation, these structural features of genuine movement participation may provide both analytical tools for researchers and strategic guidance for activists navigating an increasingly complex digital landscape. By identifying specific network metrics that predict commitment, we provide both theoretical insights into collective identity formation and practical tools for understanding-and potentially safeguarding-the dynamics of authentic digital social movements.

References

1. Ahmed, S., Jaidka, K., Cho, J.: Tweeting India's Nirbhaya protest: a study of emotional dynamics in an online social movement. Soc. Mov. Stud. **16**(4), 447–465 (2017)
2. Bail, C.A., et al.: Exposure to opposing views on social media can increase political polarization. Proc. Natl. Acad. Sci. **115**(37), 9216–9221 (2018)
3. Benkler, Y., Faris, R., Roberts, H.: Network propaganda: Manipulation, disinformation, and radicalization in American politics. Oxford University Press (2018)
4. Bennett, W.L., Segerberg, A.: The logic of connective action: digital media and the personalization of contentious politics. Inform., Commun. Society **15**(5), 739–768 (2012)
5. Bradshaw, S., Howard, P.N.: The global organization of social media disinformation campaigns. J. Int. Affairs **71**(1.5), 23–32 (2018)
6. Broniatowski, D.A., et al.: Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. Am. J. Public Health **108**(10), 1378–1384 (2018)
7. Bruns, A., Highfield, T., Burgess, J.: The Arab spring and social media audiences: English and Arabic twitter users and their networks. Am. Behav. Sci. **57**(7), 871–898 (2013)
8. Burns, A., Eltham, B.: Twitter free Iran: An evaluation of twitter's role in public diplomacy and information operations in Iran's 2009 election crisis (2009)
9. Castells, M.: Networks of outrage and hope: Social movements in the Internet age. John Wiley & Sons (2015)
10. Conover, M.D., Ferrara, E., Menczer, F., Flammini, A.: The digital evolution of occupy wall street. PLoS ONE **8**(5), e64679 (2013)
11. De Choudhury, M., Jhaver, S., Sugar, B., Weber, I.: Social media participation in an activist movement for racial equality. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 10 (2016)
12. Della Porta, D., Mattoni, A.: Social movements. Wiley Online Library (1999)
13. Earl, J., McKee Hurwitz, H., Mejia Mesinas, A., Tolan, M., Arlotti, A.: This protest will be tweeted: Twitter and protest policing during the Pittsburgh g20. Inform., Commun. Society **16**(4), 459–478 (2013)
14. Eltantawy, N., Wiest, J.B.: The Arab spring| social media in the Egyptian revolution: reconsidering resource mobilization theory. Int. J. Commun. **5**, 18 (2011)

15. Ferrara, E., Varol, O., Davis, C., Menczer, F., Flammini, A.: The rise of social bots. Commun. ACM **59**(7), 96–104 (2016)
16. Freelon, D., McIlwain, C., Clark, M.: Quantifying the power and consequences of social media protest. New Media Society **20**(3), 990–1011 (2018)
17. Giglietto, F., Righetti, N., Rossi, L., Marino, G.: Coordinated link sharing behavior as a signal to surface sources of problematic information on facebook. In: International Conference on Social Media and Society, pp. 85–91 (2020)
18. González-Bailón, S., Borge-Holthoefer, J., Moreno, Y.: Broadcasters and hidden influentials in online protest diffusion. Am. Behav. Sci. **57**(7), 943–965 (2013)
19. Goode, B.J., Krishnan, S., Roan, M., Ramakrishnan, N.: Pricing a protest: forecasting the dynamics of civil unrest activity in social media. PLoS ONE **10**(10), e0139911 (2015)
20. Harlow, S.: Social media and social movements: Facebook and an online Guatemalan justice movement that moved offline. New Media Society **14**(2), 225–243 (2012)
21. Hunt, S.A., Benford, R.D.: Collective identity, solidarity, and commitment. The Blackwell companion to social movements **433**(57) (2004)
22. Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., Westwood, S.J.: The origins and consequences of affective polarization in the united states. Annu. Rev. Polit. Sci. **22**(1), 129–146 (2019)
23. Jasper, J.M.: The emotions of protest: affective and reactive emotions in and around social movements. In: Sociological forum, vol. 13, pp. 397–424. Springer (1998)
24. Jasper, J.M.: Emotions and social movements: twenty years of theory and research. Ann. Rev. Sociol. **37**, 285–303 (2011)
25. Jost, J.T., et al.: How social media facilitates political protest: Information, motivation, and social networks. Polit. Psychol. **39**, 85–118 (2018)
26. Lazer, D.M., et al.: The science of fake news. Science **359**(6380), 1094–1096 (2018)
27. Lotan, G., Graeff, E., Ananny, M., Gaffney, D., Pearce, I., et al.: The Arab spring| the revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions. Int. J. Commun. **5**, 31 (2011)
28. Luceri, L., Deb, A., Giordano, S., Ferrara, E.: Evolution of bot and human behavior during elections. First Monday (2019)
29. Mendelsohn, J., Vijan, M., Card, D., Budak, C.: Framing social movements on social media: Unpacking diagnostic, prognostic, and motivational strategies. arXiv preprint arXiv:2406.13820 (2024)
30. Meraz, S., Papacharissi, Z.: Networked gatekeeping and networked framing on# Egypt. Int. J. Press/Politics **18**(2), 138–166 (2013)
31. Milan, S.: From social movements to cloud protesting: the evolution of collective identity. Inform. Commun. Society **18**(8), 887–900 (2015)
32. Mohammad, S.M., Kiritchenko, S., Zhu, X.: NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets. arXiv preprint arXiv:1308.6242 (2013)
33. Naskar, D., Singh, S.R., Kumar, D., Nandi, S., Rivaherrera, E.O.d.l.: Emotion dynamics of public opinions on twitter. ACM Trans. Inform. Syst. (TOIS) **38**(2), 1–24 (2020)
34. Ng, L., Carley, K.M.: A global comparison of social media bot and human characteristics. Sci. Rep. **15**(1), 10973 (2025)
35. Nizzoli, L., Tardelli, S., Avvenuti, M., Cresci, S., Tesconi, M.: Coordinated behavior on social media in 2019 uk general election. In: Proceedings of the International AAAI Conference on Web and Social Media. vol. 15, pp. 443–454 (2021)

36. Ozduzen, O., McGarry, A.: Digital traces of "twitter revolutions": Resistance, polarization, and surveillance via contested images and texts of occupy gezi. Int. J. Commun. **14**, 21 (2020)
37. Pennebaker, J.W., Booth, R.J., Francis, M.E.: Linguistic inquiry and word count: Liwc [computer software]. Austin, TX: liwc. net (2007)
38. Ransan-Cooper, H., A. Ercan, S., Duus, S.: When anger meets joy: how emotions mobilise and sustain the anti-coal seam gas movement in regional australia. Social Movement Stud. **17**(6), 635–657 (2018)
39. Sæbø, Ø., Federici, T., Braccini, A.M.: Combining social media affordances for organising collective action. Inform. Syst. J. (2020)
40. Sap, M., et al.: Developing age and gender predictive lexica over social media. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, pp. 1146–1151. Citeseer (2014)
41. Starbird, K., Arif, A., Wilson, T.: Disinformation as collaborative work: surfacing the participatory nature of strategic information operations. Proc. ACM Human-Comput. Interact. **3**(CSCW), 1–26 (2019)
42. Tremayne, M.: Anatomy of protest in the digital era: a network analysis of twitter and occupy wall street. Soc. Mov. Stud. **13**(1), 110–126 (2014)
43. Tucker, J.A., et al.: Social media, political polarization, and political disinformation: A review of the scientific literature. In: Political Polarization, and Political Disinformation: A Review of the Scientific Literature (March 19, 2018) (2018)
44. Tufekci, Z.: Twitter and tear gas: The power and fragility of networked protest. Yale University Press (2017)
45. Tufekci, Z., Wilson, C.: Social media and the decision to participate in political protest: Observations from Tahrir square. J. Commun. **62**(2), 363–379 (2012)
46. Valenzuela, S., Arriagada, A., Scherman, A.: The social media basis of youth protest behavior: the case of Chile. J. Commun. **62**(2), 299–314 (2012)
47. Van Stekelenburg, J., Klandermans, B.: The social psychology of protest. Curr. Sociol. **61**(5–6), 886–905 (2013)
48. Varol, O., Ferrara, E., Davis, C., Menczer, F., Flammini, A.: Online human-bot interactions: detection, estimation, and characterization. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 11, pp. 280–289 (2017)
49. Zellers, R., et al.: Defending against neural fake news. Adv. Neural Inform. Process. Syst. **32** (2019)