From Posts to Pavement, or Vice Versa? The Dynamic Interplay between Online Activism and Offline Confrontations

Muheng Yan[1], **Amy Yunyu Chiang**[2], **Yu-Ru Lin**[1]
[1] School of Computing and Information, University of Pittsburgh
[2] University of California, San Francisco
{muheng.yan, yurulin}@pitt.edu, {yunyu.chiang}@ucsf.edu

Abstract

This study examines how the relationship between social media discourse and offline confrontations in social movements, focusing on the "Black Lives Matter" (BLM) protests following George Floyd's death in 2020. While social media's role in facilitating social movements is well-documented, its relationship with offline confrontations remains understudied. To bridge this gap, we curate a dataset comprising 108,443 Facebook posts and 1,406 offline BLM protest events. Our analysis categorizes online media framing into "consonance" (alignment) and "dissonance" (misalignment) with the perspectives of different involved parties. Our findings indicate a reciprocal relationship between online activism support and offline confrontational occurrences. Online support for the BLM, in particular, was associated with less property damage and fewer confrontational protests in the days that followed. Conversely, offline confrontations amplified online support for the protesters. By illuminating this dynamic, we highlight the multifaceted influence of social media on social movements. Not only does it serve as a platform for information dissemination and mobilization, but it also plays a pivotal role in shaping public discourse about offline confrontations.

1 Introduction

Over the last decade, social media has revolutionized political and social activism, influencing movements such as the Arab Spring, "Black Lives Matter" (BLM), and #MeToo. The intricate relationship between online discourse and real-world events is yet to be fully understood. As events escalate into confrontations, like the BLM protests of 2020, it becomes crucial to understand whether online discussions amplify or mitigate such situations, guiding potential strategies to manage or prevent such escalations. Although traditional media has historically influenced public opinion and confrontations, as seen in Germany's right-wing violence (Koopmans and Olzak 2004), the decentralization of social media sets it apart. Much research has explored online and offline activism (Brady et al. 2021; De Choudhury et al. 2016), primarily focusing on predicting protest occurrences. Still, the linkage between online activism and specific offline events, such as confrontations, is understudied. Our study aims to bridge this gap, offering empirical insights into the correlation between online narratives and offline confrontations.

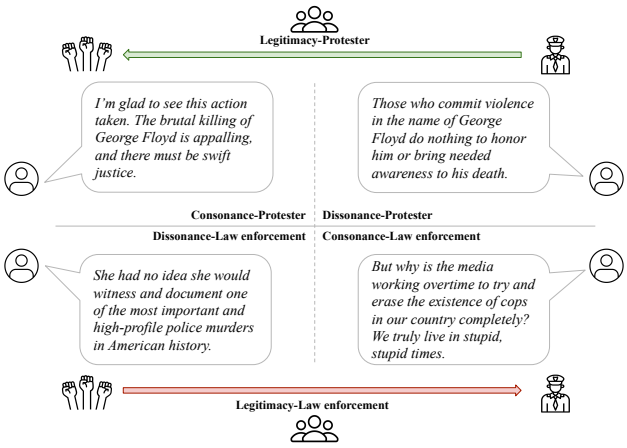

Figure 1: Example Facebook Posts of *consonance* and *dissonance*. The figure shows four example posts that are *Consonance-P*, *Dissonance-P*, *Consonance-L*, and *Dissonance-L*. From the third actors reactions to the *consonance* and *dissonance*, we then derive *legitimacy* toward each of the protester and police groups.

This study evaluates the link between online resonance and offline confrontations during the BLM movements, using data of the protests following George Floyd's death. We explore how social media discourse may predict offline confrontations and vice-versa, particularly focusing on reactions from Facebook users towards BLM protesters and the police. Recognizing the multifaceted nature of offline confrontations, our primary focus is on police and protesters – the central figures in the events. We hypothesize that the online framing can influence confrontational actions.

Inspired by the resonance theory proposed from qualitative analysis by Koopmans et al. (Koopmans and Olzak 2004), we design a computational framework to quantify the bidirectional relationship between online *resonance* (consisting *consonance*, reflecting support for a side and *dissonance*, indicating the opposition) and offline confrontation. As presented in Figure 1, resonance describes a post's inherent support or opposition, whereas *legitimacy*, derived based

on *consonance* and *dissonance*, measures third-party agreement with the post. Our analysis seeks to understand how online resonance and legitimacy correlate with offline confrontations, measured by confrontational event frequency and property damage. Distinct from the general "public opinion" as measured by surveys (Horowitz et al. 2023; Joseph et al. 2021), online resonance is immediately visible to the public, thereby potentially influencing the movement's trajectory. This study uses Facebook data to gauge online resonance and to expand the theory into a computational framework, allowing the assessment of how online resonance is related to offline confrontation.

The key challenges to study the relationship between resonance, legitimacy, and confrontation lie in (1) the lack of empirical resonance measurement based on social media data in the context of protests, and (2) the statistical uncertainty in analysis for studying the online-offline reciprocal relationship. This study tackles both the challenges that results in the following contributions:

- We introduce the novel concept of online resonance. We establish an annotation scheme and a ground-truth dataset of online resonance using data collected from Facebook. In conjunction with data documenting protest activities across the U.S. during the BLM protests triggered by George Floyd's death, we present the first large-scale study of online resonance[1].
- We leverage a few-shot machine learning model to measure online resonance at scale. Under a variety of stringent conditions and with relatively small label sets, our results demonstrate a performance improvement of up to 18% in classifying Facebook users' attitudes expressed in their posts, compared to baselines.
- We propose an analysis framework to study the reciprocal relationship between resonance and confrontation, which includes (a) the measurement of resonance at the collective level, and (b) the time-series analysis to explore the resonance-confrontation relationship and directionality.
- We present new empirical findings that disentangle the competing theories about the relationship between online resonance and offline confrontation risks. Rather than a straightforward positive or negative feedback loop between the two, our analysis reveals a reciprocal relationship that involves both positive and negative associations, where more online support correlates with a lower level of confrontation risk, whereas a higher level of confrontation risk likely is associated with more online support.

While our study doesn't confirm causality, it highlights a predictive link between social media discourse and confrontations. Our findings offer insights for confrontation escalations and suggest that understanding online reactions could help prevent such confrontations during protests.

2 Related Work

Confrontation and Public Discourse. Offline violence or confrontations are less effective in achieving protesters' goals and success (Chenoweth and Cunningham 2013; Huet-Vaughn 2013). Negative impacts on public support may limit success (Muñoz and Anduiza 2019), with success depending on public opinion and attitude (Mazumder 2018; Edwards and Arnon 2021). However, offline violence sometimes does result in unintended consequences when they are used to enhance the discourses of the elite based on public order maintenance (Wasow 2017); or to reinforce the blame for the use of violence by the opponent (Yan et al. 2017; Howes and Classen 2013). While little evidence suggests BLM protesters were widespread violent, the "biased social media framing" disproportionately highlighted looting, vandalism, and interpersonal conflicts, which possibly contributed to the reduced public support for the BLM movement (York 2022).

These findings highlighted the paradoxical relationship between the use of violent confrontations and public discourse: offline violence tends to harm popular support for social movements, yet social media framing of these events sometimes may increase public support of the said events, especially when violence is perceived as justified in certain circumstances (Koopmans and Olzak 2004; Shuman et al. 2022). This complexity suggests that, while physical confrontations can initially alienate potential supporters, the narrative constructed around these events in online spaces can alter public perception. This interaction between offline actions and online discourse underscores the importance of media framing in shaping public opinion about social movements, indicating a nuanced yet under-explored relationship between real-world events and their digital representations. This research seeks to examine how online social media affects public attitudes, especially willingness to engage in offline confrontations, and vice versa.

Social Media and Offline Collective Action. Social media are essential for understanding public opinion and online activism due to its ability to facilitate offline collective action (Ertugrul et al. 2019; Brady et al. 2021). Because of the effective information exchange on social media, it is efficient to coordinate gatherings, promote agendas, and report activities in a variety of movements, such as Arab Spring and the "Black Lives Matter" (BLM) movements (Wei et al. 2020; Greijdanus et al. 2020). Social media contributes to the acceleration of processes of sense-making that involves perceptions of efficacy, emotions, and social identity (McGarty et al. 2014; Drury and Reicher 2005). For instance, social media boost the formation of collective identities (Tajfel et al. 1979) that further results in feelings of obligation to engage in social movements (Sturmer and Simon 2004), evident in the BLM movements. The use of social media tends to facilitate such social psychological antecedents and, subsequently, protest participation. Some argue that social media also gives rise to "slacktivism", a disconnect between awareness, support, and social media participation (Cabrera et al. 2017; Yan, Lin, and Chung 2022), while others suggest social media can lead to offline collective action such as protests (Wilkins et al. 2019). Our research aims to bridge the gap in understanding the reciprocal influence between online activism and offline collective action, an aspect that remains insufficiently explored in current literature.

[1]https://github.com/picsolab/OnlineResonance-OfflineConfrontation-Dataset

Attitude Detection with Few-Shot Learning. Social media *consonance* and *dissonance* are collective online opinions towards confronting groups. While Sentiment Analysis models have been robust in analyzing political and social issues (Elghazaly 2016), they may not suffice for measuring nuanced resonance aspects. These require both sentiment polarity and targets, linking it to Stance Detection in NLP (Küçük and Can 2020; Yan et al. 2020). While there are efforts in stance detection, such as predicting attitudes towards specific topics (Darwish et al. 2020), or predicting the target and attitude (Dey et al. 2017), framing label prediction poses challenges that arise from similar frames possessing variable sentiments and the scarcity of frame-labeled samples, making traditional ML training problematic.

The emergence of deep learning models like BERT (Devlin et al. 2018; Liu et al. 2019) has revolutionized this domain. With pre-training on vast text corpora, these models have demonstrated efficacy in NLP tasks including Stance Detection. However, transiting the richness of linguistic patterns learned during pre-training to classification labels remains a challenge. This is evident when these labels, such as binary outcomes, do not intuitively map to "negative/positive." The innovative approach of few-shot learning offers a solution by converting numerical classification labels into cloze questions (Schick and Schütze 2020). This transformative method allows models to achieve efficient and accurate performance with significantly less labeled data.

3 Theoretical Framework

We define offline confrontations in social movements as those that are: (a) non-peaceful protests or gatherings. (Thomas et al. 2014), and (2) reportedly directed at either the protesters or the police that included some forms of confrontation or the types of collective actions that are often referred to as "non-normative',' or "disruptive", "nonviolent" collective action in the literature. (Thomas et al. 2014; Chiang 2021). Following the definition in previous studies (Thomas et al. 2014; Chiang 2021), we consider protest events that include physical and non-physical damage or harm, such as arrests, property damages, and injuries, to be *confrontational*. In this study, we characterize the intensity of confrontations by (i) the total number of confrontational events (such as arrests and injuries) and (ii) protester-initiated property damage as surrogates for the magnitude of confrontations, as confrontations are likely to involve property destruction or violence incidents.

For the analysis of online *resonance*, we examine the concepts of *consonance* and *legitimacy*, first introduced by Koopmans and Olzak (Koopmans and Olzak 2004) to analyze the relation between mainstream media and movement confrontation. In our context, we define *consonance* as social media authors expressing support for a social movement's actions or demands. On the contrary, *dissonance* reflects the social media authors' rejection of the movement's demands and actions. Both the concepts can have directions toward both sides of the confrontations – one can support or oppose either protesters or the police – and our focus is on the protesters. *Legitimacy* is the degree to which reactions by third actors in the public sphere support an actor's claims more than they reject them, where in this study, an actor who made a claim refers to the author of a social media post, and third actors are social media users who viewed and responded to the post. *Legitimacy* can vary independently of consonance because it signals how *widely* the claims and support from the protesters or the police have spread in public. In Figure 1, we show four Facebook posts that exemplify either *consonance* or *dissonance* towards either the protesters or the police, as well as the *legitimacy*, which is measured based on aggregating the reactions (e.g., likes) on posts expressing *consonance* or *dissonance*.

Existing literature has emphasized the role of social media in facilitating the exchange of information and the support for the protest activities, both of which are vital to the coordination of the protest and to mobilize and attract participation (Boulianne 2018). Social media attention and support are argued to be a catalyst for protest mobilization (Boulianne et al. 2020; Breuer et al. 2015). However, the relationship between social media attention and offline confrontation escalation is less understood. Prior studies suggest a positive relationship could exist between social media attention and offline confrontation, as activists use offline confrontation as a tactic to seek media attention or coverage (Koopmans and Olzak 2004; Shuman et al. 2022); On the other hand, a negative relationship between the two also seems plausible because offline confrontations (e.g., protester-initiated confrontations) tend to weaken the *legitimacy* of protests, which could lower their online support and public sympathy while increasing the legitimacy of the use of violence by the authority (Wasow 2020). These competing theories not only differ in *signs* but also in *directions*. This work seeks empirical evidence for these competing theories by asking the following three research questions that examine the bidirectional relationship between the online *consonance* and *legitimacy* and the offline confrontations:

RQ1: How is online consonance with the BLM protesters associated with offline confrontations?

RQ2: How is online legitimacy with the BLM protesters correlated with offline confrontations?

RQ3: Do offline confrontations shape online consonance or legitimacy as well?

Confrontations often emerge when peaceful alternatives do not achieve the desired outcome. For **RQ1**, we anticipate that greater online *consonance* with protesters would correlate with reduced offline confrontations. Historical data suggests that confrontational tactics can diminish public support, and alienate potential sympathizers, making it challenging for people to identify with or justify the movement's actions (Muñoz and Anduiza 2019). On the other hand, people who engage in political violence may do so as they perceive that the political system is ineffective and that extreme methods are the only recourse (Spears 2010).

For **RQ2**, increased online *legitimacy* tied to the protesters might reduce the impetus for violent confrontations offline. While consonance and legitimacy are related, they exert influence differently. For instance, strong consonance with police might either undermine the protest's legitimacy or render its claims controversial, depriving it of media and public attention. In such scenarios, movements might employ

drastic measures, like confrontations, to reclaim attention. Evidence suggests that protester-initiated confrontations can shift the narrative from sympathy for the movement to the necessity of societal control by authorities (Wasow 2020).

Finally, for **RQ3**, research indicates that those previously engaged in offline activism might pivot to online methods, an effect termed the "spillover hypothesis" (Kim et al. 2017). This effect is pronounced among young adults and through interconnected social networks (Greijdanus et al. 2020). Additionally, there is evidence of more protests in cities with democratic leanings (Williamson et al. 2018). Our analysis, using city-day data, will consider these variables alongside socio-demographic factors such as political orientation and racial compositions in cities with active BLM movements.

4 Resonance-Confrontation Dataset

We summarizes the steps we take to create the dataset and conduct the analysis in five steps including a) Data Collection, b) Attitude Annotation, c(Attitude Label Augmentation via Few-Shot Learning, d) Building Resonance-Confrontation Dataset, and e) Analysis to Answer the RQs. The steps are illustrated in Figure 6 in the Appendix. We start by developing the *Resonance Confrontation Dataset*.

4.1 Data Collection

Social Media Data. We utilize CrowdTangle to gather Facebook posts about the George Floyd Protest from May 26th, 2020, a day after his death, until June 14th, 2020, using the keyword "George Floyd." CrowdTangle's APIs allow retrieval of historical data from public Pages or Groups, ensuring a comprehensive collection without sampling biases. During this 19-day span, we collect 267,522 posts.

To relate online data with offline events, we focus on city-level geo-location, i.e., associating online posts with real-world events in specific cities. If posts lacked geo-tags in their metadata, we infer locations using Named Entity Recognition, taking the first identified location as the post's subject. Importantly, we aim to identify posts' referring locations, not their origin. This approach captures the online discussion about a specific place rather than where the discussion began. It is more relevant to determine what the post's content is about than where it originated from. If a location is not mentioned or inferred from context, determining support or opposition in the post becomes ambiguous. Of the collected posts, 108,443 had geo-locations, either from metadata or our inferences. Figure 2 presents the online and offline activities in our dataset, both aggregated and daily.

Protest Data. We collect offline protest information from the crowd-counting-consortium (CCC) dataset[2]. The CCC dataset has gathered information on all types of protests in the United States since 2017. We filter the protests by the date range, and get 1,406 events across 58 cities in the US. Each protest in the CCC dataset is recorded with violence tolls, including the number of people injured, the number of police injured, the number of people arrested, and the value of property damage. We consider events that have been reported with **any** protester injured, police injured, or people arrested as *confrontational* events and others as *non-confrontational*. Figure 2 plots the number of confrontational and non-confrontational activities by date.

Figure 2: Online and offline activities. (A) The map shows the US cities' activities captured from the Facebook posts. Each circle corresponds to a city, with size indicating the total number of posts during the study period (logarithmic scale). The top 20 cities with the most posts are labeled on the map. (B) The heatmap shows the day-by-day online post and offline protest activities for the top 20 cities, with size indicating the daily post count (logarithmic scale) and color indicating the daily protest count.

As outlined in Section 4.5, our location matching between Facebook and CCC data identifies 57 cities, accounting for 73% of BLM protests recorded in the CCC data. It is crucial to recognize that our Facebook-CCC dataset is not representative of the U.S. population, thereby limiting the direct applicability of our results to broader offline scenarios. However, this study's primary objective is to determine the relationship between online resonance and offline confrontations. Given the widespread use of Facebook in the United States, as supported by a PEW study (Auxier and Anderson

[2]https://crowdcounting.org

2021), our method provides unique insights into the dynamics of online and offline interactions.

4.2 Identifying Resonance in Facebook Posts

Public perceptions form the resonance. Following the death of George Floyd in 2020, the public support diversified toward the two opposing groups: the BLM protesters and the police. Unlike other protests, the "police" is highly related to the "police violence/brutality" that the demonstrations protested against followed by the killing of George Floyd by a police. This protest framing intensified the confrontation between the protesters and the police during the demonstration period (Horne 2022), and has made the BLM movement unique to others – it grants us a chance to study the public's resonance framing beyond a binary setting of only consonance or dissonance toward a certain target. As shown in Figure 1, there are two targets (the police or the protesters) by two resonance values (consonance or dissonance), resulting in four classes: *Consonance-P*, *Dissonance-P*, *Consonance-L*, and *Dissonance-L*, in which the "P" refers to the **P**rotesters and the "L" refers to the police, namely **L**aw Enforcers. Table 4 in the Appendix shows all acronyms used in this study and their descriptions. Furthermore, for each of the `P` and `L`, we are able to investigate the audiences' reactions (likes, re-shares, and loves) to the resonance as *Legitimacy-P* and *Legitimacy-L*, indicate the extent to which the audiences in general lean toward *consonance* (+) or *dissonance* (-). The absolute values of the constructs reflect the degree to which the audiences agree with the consonance or dissonance voices.

4.3 Annotating Attitudes in Facebook Posts

To measure the resonance at the collective level, we first need to identify the *attitudes* in each post. We develop a set of human annotated data as ground-truth, and employ computational tools to scale up the post labeling process.

Annotation Overview. To quantify the sentiment of Facebook posts, we first obtain human annotations as our ground-truth. Two challenges emerged: (1) some collected posts related to the protests lacked clear attitudes regarding protesters or police, and (2) users sometimes reshare posts but added comments with differing sentiments. To address these, we introduce neutral categories: "not targeting any specific group" and "no explicit sentiment displayed." Additionally, annotators were instructed to label stance in the original post ("author attitude") and reshared content ("cited attitude") separately. These labels comprise seven choices, including three sentiments (favor, against, neutral) with two targets (police, protester), and a "no target" option.

Data Sampling for Annotation. We adopt a sampling strategy to minimize the noise introduced later in the augmented labels by machine learning algorithms (ref Section 4.4). Specifically, we rank the posts by their interactions received, by city and by day, and prioritize the high-interacted posts to human annotation. We take the 2.5% top-interacted of the Facebook posts (which resulted in total 857 posts) per day/city for the annotators to label, and use the machine learning algorithm described below in Section 4.4 to infer the label for the rests.

(a)

Author's Attitude	A1	A2	A3
A1	/	0.515	0.623
A2	*0.908*	/	0.545
A3	*0.885*	*0.884*	/

(b)

Cited Attitude	A1	A2	A3
A1	/	0.719	0.528
A2	*0.902*	/	0.498
A3	*0.867*	*0.867*	/

Table 1: The inter-coder reliability between the annotators. The Cohen's Kappa (Upper Triangles) and Gwet's AC1 (Lower Triangles) scores between any pair of annotators for the annotation (three categories) on **(a)** post authors' attitudes, and **(b)** cited attitudes, in 570 samples.

Annotators and Annotation Scheme. Three annotators followed the aforementioned annotation guideline to create ground-truth labels All three annotators are native English speakers. We train the annotators with 287 samples, and then pair the three annotators with each other to annotate the other 570 samples. If there is a disagreement, the other annotator is involved as a tie-breaker.

Annotation Evaluation. We calculate inter-rater agreement scores for all annotator pairs for both "authors' attitude" and "cited attitude" questions. In Table 1, we present a detailed analysis of pair-wise agreements for both "author's attitude.. and "cited attitude" variables. This analysis employs two reliability metrics – including Cohen's Kappa and Gwet's AC1 (Gwet 2001) – to gauge the consistency of our annotators' evaluations. The reason for employing Gwet's AC1, in addition to Cohen's Kappa, stems from the latter's vulnerability to skewed label distributions. In our dataset, such an imbalance is evident, as demonstrated by the predominance of the "No Target" label, which constitutes over 60% of all labels, as shown in Table 2. The results of our coding process indicate a commendable level of agreement among annotators. Specifically, the Cohen's Kappa scores range from "Moderate" to "Substantial," signifying a reliable level of consistency in the annotators' assessments. Furthermore, the Gwet's AC1 scores fall into the "Very good" category, underscoring the robustness of the agreement across our annotating pairs, exceeding or resembling the agreement levels reported in other published studies (Maloney et al. 2023).

We then generate the overall attitude for each Facebook post from the authors' and cited attitude by the following logic: if there is an "authors attitude", it overrides the "cited attitude" and becomes the post's attitude; otherwise, if there is no target, or the attitude is "neutral" in the "author attitude", the post is annotated as the "cited attitude". The heuristic is that if a post author *does not* explicitly express an attitude, the post is only amplifying the attitude in the cited materials. Table 2 reports the number of posts under the label from the annotation results.

4.4 Few-shot Attitude Prediction

Label Augmentation with NLP Model. We develop tools to label 108,443 samples in our corpus based on 857 labeled ones for our task of predicting attitude (support or oppose) toward a target (protesters or police), similar to "stance prediction" in NLP (Küçük and Can 2020). We adopt this prob-

	#Posts	Ratio	#Training Samples		
			Low	Mid	High
Support Protesters	118	13.8%	12	30	47
Against Police	29	3.4%	3	7	12
Support Police	5	0.6%	1	1	2
Against Protesters	66	7.7%	7	17	26
Neutral Protesters	100	11.7%	10	25	40
Neutral Police	23	2.7%	2	6	9
No Target	516	60.2%	33	86	136
Total	857	100.0%	84	172	273

Table 2: The number of posts per category from annotation. The last three columns reports the training sample size of each label category to be used in the Attitude Prediction task. The sampling method is reported in Section 4.4.

lem scheme for augmenting attitude labels. Our study uses data with a ratio of labeled to unlabeled samples under 1%. For this scarce resource setting in which conventional methods – even the fine-tuning of deep NLP models – would fail, we leverage PET (Schick and Schütze 2020), a state-of-the-art few-shots model to achieve the best performances. PET requires fewer samples to classify stance than the other approach by associating the classification labels with semantic meanings in "cloze" questions (Schick and Schütze 2020). In this study, we use the ROBERTA pretrained model (Liu et al. 2019) to initialize both the PET model and a baseline BERT fine-tuning model in our experiments.

A cloze question is a sentence-completion task where machine learning models predict missing words based on the given context. Instead of a standard classification, such as predicting a zero or one outcome, the model might fill in a blank in "This sample is _____." with either "positive" or "negative." Since large-scale pre-trained models are often trained on sentence completion tasks, converting a classification task to a cloze format allows for utilizing models' pre-existing knowledge, potentially improving accuracy without extensive training for specific classification tasks.

We create a two-step cloze-question style classification to augment the attitude labels for the social media posts:

Step 1. We ask the model to predict the target of attitude, if it exists. We create the following cloze question:

"I am commenting on _____. [POST]",

where the [POST] is replaced by the content of each post sample, and the blank is one of the three predictions outputs: *the protesters*, *the police*, and *nothing*.

Step 2. We use the model's prediction in the first step as a prompt to predict the stance value. The cloze question is

"I am _____ [TARGET]. [POST]",

where the [POST] is replaced by the content of each post sample, the [TARGET] is the output of Step 1, which is one of the following: *the police*, *the protesters*, or none (without any target keywords), and the blank is one of three prediction output: *positive*, *negative*, and *neutral*.

Experiment Setting. We compare the few-shot learning model to two other baseline models: a BERT model with a dense layer for fine-tuning and an n-gram model with a Random Forest classifier.

To address the imbalanced label distribution (Table 2), a down-sampling procedure is applied. For each label, except "No Target", $x\%$ is randomly chosen for both training and development sets, leaving the remaining (1 - $2x\%$) for testing. An equal number of samples is taken from the "No Target" label for consistent representation. Given $x = 10$, 10% instances from each category (i.e., 12, 3, 1, 7, 10, and 2), 33 in total, will be sampled, and 33 "No Target" instances will also be sampled. The experiment has three conditions: low-resource (84 samples), mid-resource (172 samples), and high-resource (273 samples), as shown in Table 2. Even the high-resource condition contains a small sample size.

Two prediction tasks are formulated: one using seven fine-grained labels and another using three coarse-grained labels. In the latter, close attitudes (i.e., *Consonance-P* and *Dissonance-L* together; *Consonance-L* and *Dissonance-P* together; and neutral categories together) are merged. Employing seven fine-grained labels is indicative of a comprehensive and nuanced understanding of the attitudes into specific categories. Conversely, the second task, utilizing three coarse-grained labels, adopts a more aggregated approach. This is due to the sparsity of both *Consonance-L* and *Dissonance-L* labels in the dataset. Merging the labels allows us to discern the predominant resonance of the confrontation between police and the protesters.

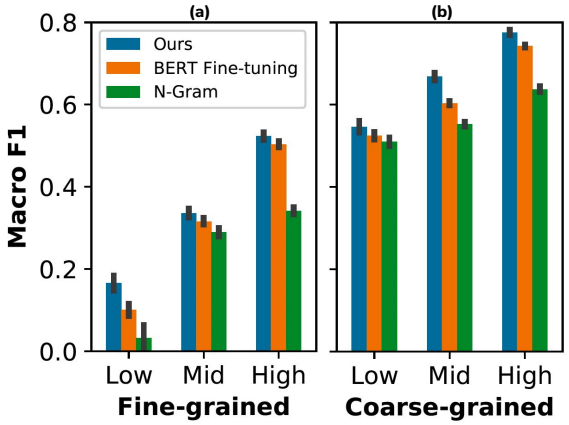

Figure 3: Performance of Attitude Predictions tasks. Weighted F1 scores for the compared models on different resource conditions, with **(a)** seven labels, and **(b)** three coarse-grained labels. Detailed performances are reported in Appendix A.

Evaluation. Each model is evaluated per label and via a weighted F1 score considering label distributions. Figure 3 shows the weighted F1 scores across all settings. Our few-shot learning model excels with limited training resources but sees diminishing gains with increased resources. In contrast, the BERT model shows significant performance improvements as training resources expand, which is consistent with previous research (Schick and Schütze 2020).

4.5 Analysis of Resonance

Combining the predicted labels of all 108,443 posts and the offline protest events from the CCC dataset, we create the

Resonance Confrontation Dataset from these data in this sub-section, and report the trend of online resonance.

Geo-temporal Grouping. We first group the Facebook posts and the offline demonstrations by city and by day. It enables us to better fine-grain the resonance and confrontational violence measurement, and also better associate them according to the geo-temporal relatedness. After grouping both data by city and day, we discard city data with too scarce data points. Only cities: 1) having at least one protest happened during the 19-day period, and 2) having posts in at least 14 days. After this filtering process, we have in total 48,817 posts across 57 cities in the U.S., for the 19-day protest period. These 57 cities account for 73% of the total number of BLM protests captured in the CCC data. They cover 38 out of the 50 largest cities in the US; the cumulative populations of these cities cover nearly half (49%) of the U.S. population [3].

Measuring the Values for Resonance Constructs. We then calculate the values for the six constructs defined earlier in Section 4.2: *Consonance-P*, *Dissonance-P*, *Consonance-L*, *Dissonance-L*, *Legitimacy-P*, and *Legitimacy-L*. In specific, we measure the `Consonance` (for both protesters `P` and law enforcers `L`) as the fraction of posts with *positive* attitude toward the target, per city per day. The `Dissonance` is calculated similarly with the positive changed into *negative*. Due to the sparsity of *Dissonance-L* and *Consonance-L*, we focus on *Dissonance-P* and *Consonance-P* in the following analysis presented in Section 5. Lastly, we measure the `Legitimacy` as **the proportion of interactions** (likes, reshares, and loves) of the "positive-[TARGET]" posts, **minus the proportion of interactions** of the "negative-[TARGET]" posts. The [TARGET] refers to either `P` or `L`.

5 Analysis Methods

Autoregressive Distributed Lag Framework. We employ an Autoregressive Distributed Lag (ARDL) framework to examine the relationship between online resonance and offline confrontation. ARDL is a widely used time-series analysis for the temporal dependency between variables. It incorporates a lag structure from both the endogenous variable and the exogenous variables. The full specification of an ARDL model includes trend and seasonal components, autoregressive terms, exogenous regressors, and other fixed regressors without the distributed lag structure. Due to the short observed time periods from our data, we omit the trend and seasonal components. And, given the skewed distribution in the online and offline observations across cities, we consider a logistic model between the regressors and a binarized outcome than a linear model. Our specification is:

$$Y_t = f(\alpha_0 + \sum_{p=1}^{P} \phi_p Y_{t-p} + \sum_{k=1}^{M} \sum_{j=0}^{Q} \beta_{k,j} X_{k,t-j} + \sum_{l=1}^{L} \gamma_l Z_l^{(t)}) + \epsilon_t, \quad (1)$$

where α_0 is a constant, ϕ_p and $\beta_{k,j}$ are the coefficients of an autoregressive term with lag p, and of an exogenous (independent) variable $X_{k,t-j}$ with lag j, respectively. Z_l is a fixed regressor, ϵ_t is the binomial error term corresponding to the logistic function $f(\cdot)$.

City-level covariates and time-dependent control variables. Inspired by previous research showing the effect of sociopolitical and population characteristics on protests or confrontational actions (Steinert-Threlkeld et al. 2022), we include a set of city-level covariates. We use the party of the state legislatures as a proxy for political leanings and capture the city's population characteristics with population and density. In addition, since cities' online and offline activities can be influenced by national activities, we include a control variable of the global (national) social media activities, captured as the sum of all protest-related Facebook posts in our dataset. The unit of analysis is a city, with daily temporal observations recorded. The outcome variable Y_t is considered in two cases: (a) as an offline confrontation index, such as the number of confrontational protests and property damage, and (b) as an online measure, such as consonance or legitimacy. Using the median as a threshold, the outcome variable is transformed into a binary value. When the outcome variable is an offline index, an online measure is used as the independent variable, and vice versa. The independent variables that are shared between the two cases are the city-level covariates and the global, time-dependent control variable, as previously described. In order to account for short-term fluctuations and improve the estimates' reliability, we employ a three-day rolling average on both outcomes (left-aligned) and time-varying independent variables (right-aligned) to ensure the predicted values of the outcome variables occur strictly after those of the independent variables.

Structural Equation Modelling for Mutually Related Outcomes. The ARDL framework allows us to examine the online-offline relationship among variables separately via dynamic single-equation regressions. However, the online and offline variables may be mutually correlated. For example, the online activities may be influenced by the offline outcome in the past that is related to an even earlier outline outcome. In this case, it is interesting to test whether there is a reciprocal relationship between the two types of outcomes. We use structural equation modeling (SEM) (Hox and Bechger 1998), a general statistical framework to test the structure of the two related outcomes and other independent variables. The main difference between two single-equation regressions and the SEM specification is that the errors of the predictor variables, as well as the errors between the two outcome variables may be correlated, and we use SEM to incorporate the covariance structures among related variables to obtain less biased estimates.

While Granger causality analysis has been extensively utilized for analyzing the relationship between time series data, its recent development has increased its utility, such as when dealing with multivariate or nonlinear time series (Shojaie and Fox 2022). Nevertheless, given the small sample sizes (each time series contains no more than 19 observations), it is unlikely to derive appropriate estimates even for a standard Granger causality test (Ramos and Macau 2017). To balance model complexity/assumptions and practical applicability, we therefore opt for ARDL and SEM to examine

[3]The population information is from government report at: https://www.census.gov/quickfacts/fact/table/US/PST045221

the relationship between online and offline observations.

6 Results

$RQ1$**: How is online *Consonance-P* associated with offline confrontations?** First, we examine the relationship between online *consonance with the protester side* and offline confrontations (number of confrontations and property damage). Note that *Consonance-P* is measured as the number of posts supporting the *protester side* per day/city. We use the *consonance* from $t-1$ and $t-2$ to assess its impact on each offline outcome at time t. With the ARDL models, we find that *Consonance-P* is associated with *fewer* offline confrontations: one unit change with online *Consonance-P* from time $t-1$ is associated -0.42 ($p < 0.001$) for $t-2$ change of offline confrontations, and -0.73 ($p < 0.00$) decrease in property damage at time $t-1$ as well. All of the continuous predictors are standardized. Among the covariates included in the models, population density is positively correlated with the outcome *offline confrontational* events ($\beta = 0.46, p < 0.001$). Having a Democrat governor in a city reduces the likelihood of confrontations ($\beta = -0.42, p < 0.01$). Having a less diverse population in a city is associated with fewer offline confrontations ($\beta = -0.28$ for % of the White population in the city, $\beta = -0.24$ for % of the Black population, both with $p < 0.01$). A similar pattern can be found concerning the outcome *property damage*, although most covariates fail to reach statistical significance.

The results show that the *Consonance-P* is negatively associated with offline confrontations. The more the BLM demonstrators have consonance online, the less likely we would observe protest confrontations in the future days.

$RQ2$**: How is *Legitimacy-P* correlated with offline confrontations?** *Legitimacy-P* appears to be associated with *fewer* offline confrontations: more *Legitimacy-P* is associated with fewer confrontations ($\beta = -0.04$) at time $t-1$, and ($\beta = -0.31, p < 0.00$) at time $t-2$, although the coefficient at $t-1$ fails to reach statistical significance. In a similar vein, *Legitimacy-P* is also linked with *fewer* episodes of property damage ($\beta = -0.34, p < 0.001$) at time $t-1$.

We continue to observe that the higher the population density is, the higher the likelihood of offline confrontations ($\beta = 0.44, p < 0.001$). Having a Democrat governor ($\beta = -0.38, p < 0.05$), more percentage of the White population ($\beta = -0.26, p < 0.01$), and more percentage of the Black population ($\beta = -0.23, p < 0.01$), quite to the contrary, reduces the likelihood of offline confrontations. Given the predominant fraction of the White population in the US, a higher percentage of the White population indicates a more homogeneous population in the cities, lessening the chance of large-scale BLM protests (Horowitz 2022). On the other hand, a larger Black population indicates a more diverse population in the cities. This may promote the liberal ideology which is linked to higher adaptation probability of the BLM protest agenda, and thus reduce the chance of direct confrontation in protests.

The predictive trends of *consonance* and *legitimacy* are similar: the more the BLM demonstrators get support from the online populations, the less likely we would observe confrontations in the future events. It is likely due to that the movements do not need to resort to confrontation to draw more attention from the public to get their agenda noticed.

$RQ3$**: Do offline confrontations shape online `consonance` or `legitimacy` as well?** We also consider the possibility that offline confrontations shape online social media consonance and legitimacy - i.e., how individuals and the public sphere react to offline confrontations. Results generally support this argument. For the outcome of *Consonance-P*, we find that the higher the number of population, the lower likelihood of *Consonance-P* ($\beta = -0.4$, $p < 0.001$). Moreover, it appears that the higher percentage of the White population is in a given city, the lower likelihood of *Consonance-P* ($\beta = -0.19$, $p < 0.05$) to be observed. The same does not apply when *Legitimacy-P* is the outcome.

Interestingly, offline confrontations are associated with higher online *Legitimacy-P*: ($\beta = 0.35, p < 0.001$). This leads to an interpretation that if the BLM demonstrators adopt confrontations in offline protests, they may draw more support from the public online.

Term	Consonance				Legitimacy			
	Est.	CI	EV	EV(LB)	Est.	CI	EV	EV(LB)
X_{t-2}	0.66	[0.57, 0.76]	2.41	1.95	0.73	[0.63, 0.85]	2.07	1.63
X_{t-1}	1.42	[1.22, 1.65]	2.18	1.73	1.33	[1.15, 1.54]	1.98	1.55
X_{t-1}	0.48	[0.38, 0.60]	3.57	2.72	0.71	[0.59, 0.86]	2.15	1.58
X_{t-1}	1.42	[1.21, 1.68]	2.18	1.71	1.30	[1.14, 1.52]	1.93	1.53

Table 3: Sensitivity analysis based on E-value. It suggests that the estimated effects are relatively robust to an unmeasured confounder.

Sensitivity analysis with unmeasured confounders. Despite accounting for major theoretically relevant covariates, our regression analysis may still be affected by latent factors like traditional media reporting or activities on platforms other than Facebook. To assess the robustness of our estimates against potential confounding, we employed the E-value method (VanderWeele and Ding 2017). The E-value quantifies the minimum association strength required between unobserved confounders and the variables to nullify our observed association. Table 3 presents results for consonance-confrontation and legitimacy-confrontation effects, detailing coefficient estimates, confidence intervals in OR, and derived E-values. The consonance-confrontation effect analysis yields E-values ranging from 2.18 to 3.57. An E-value of 2.18 implies that any unmeasured confounder would need a 2.18-fold risk ratio association with both consonance and confrontation, after considering all measured covariates, to negate our results. The magnitude of the E-values strengthens our findings' robustness, indicating that it is highly unlikely that unmeasured confounding can easily reduce the observed effect to the null effect.

Robustness checks with bidirectional relationships. We further experiment to confirm the bidirectional relationships discovered in regression models using structural equation modeling (SEM). The results can be found in Figure 5. In general, the results from SEM modeling provide further evi-

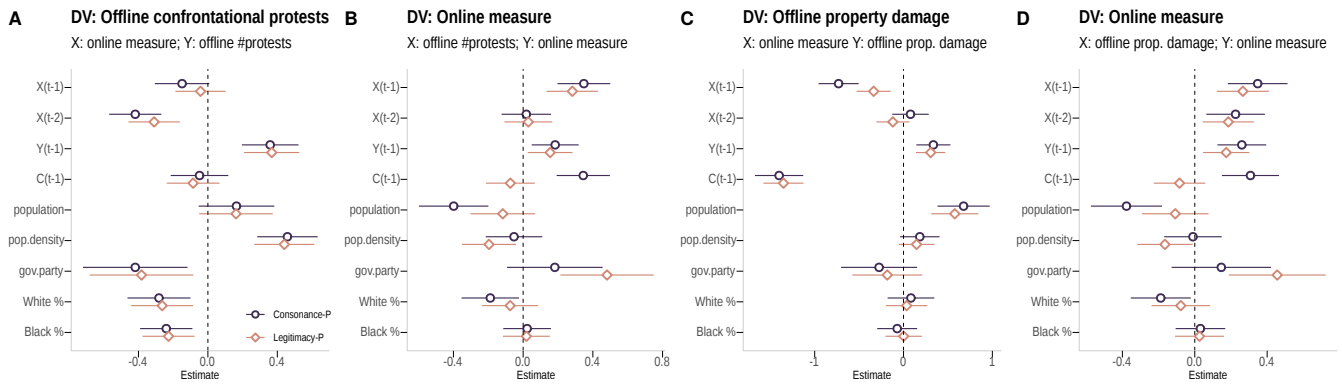

Figure 4: Predictive analysis for online-offline association using ARDL models. X, Y, C denote the main predictor(s), outcome variable, and the co-variate that controls the global online activity, respectively. The estimated effects are presented in standardized log-odds.

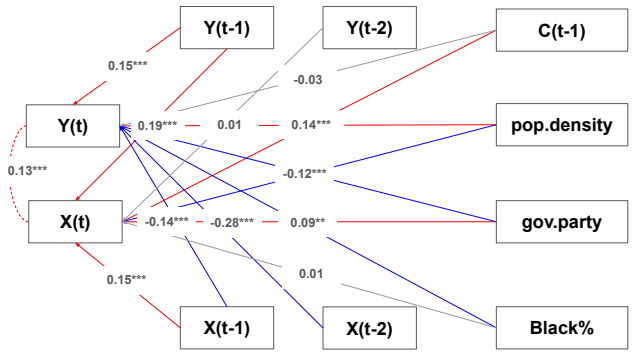

Figure 5: Predictive analysis for online-offline association using SEM. Two outcomes, $X(t)$ and $Y(t)$, are the online consonance and offline violent protests. The predictors include lagged online and offline measures, the global online activity $C(t-1)$, and the city attributes. The effect estimates are labeled on the corresponding edges, with a red/blue color indicating a significant positive/negative association. The dashed line indicates the covariance between variables.

dence for our main findings in ARDL modeling in Figure 4. First, *Consonance-P* significantly predicts *fewer* offline confrontational events: $\beta = -0.14, p < 0.001$ at time $t-1$, and $\beta = -0.28, p < 0.001$ at time $t-2$. Second, offline confrontational events at time $t-1$ are associated with more *Consonance-P* ($\beta = 0.19, p < 0.001$). Finally, regarding the covariates, having a higher percentage of the Black population in a given city during the BLM movements is associated with *fewer* offline confrontational events. Having a democrat governor results in fewer offline violent events, but more *Consonance-P*. Overall, it appears that the magnitudes of effect are more substantial when we account for covariance between the online *consonance* and offline confrontations. The findings also suggest that there exists non-trivial reciprocal dynamic relationships between online social media activities and offline confrontations.

7 Discussions

This study investigates how social media resonance is associated with the likelihood of offline confrontations and vice versa. Our empirical study contributes to the existing literature by identifying social media's online consonance and legitimacy aspects and their impacts on offline confrontations, revealing that online and offline activism may mutually correlate with each other. Our study suggests that high online consonance or legitimacy with the protesters is associated with *fewer* offline confrontational events and property damage. Reversely, offline confrontations likely *correlate* with online consonance and legitimacy to the protesters in the BLM protests. The results indicate that the online and offline link between social media and offline activism is not a straightforward positive or negative relationship as the existing literature suggests, but a reciprocal and bidirectional relationship that can be distinguished when considering the temporal order of events.

Authorities usually put emphasis on the violent aspect of protests (Brown and Mourão 2021). Our findings provide insights into de-escalation strategies. Most likely, the less legitimacy the protesters feel, the more likely they will risk using confrontational strategies. However, emphasizing the confrontational aspect would instead increase the risk of police-protest tensions. Recognizing the protesters' agenda and showing support with their sensible appeals or demands may instead lead to a more peaceful outcome. The purpose of offline protest is to gather support and generate participation. When support is observed, the need for further offline activities, especially the ones with confrontations, is significantly decreased. On the contrary, when the public backs the authorities, it can often justify the use of force and lead to an escalation of tension between the protesters and the authorities. The findings add to the body of evidence suggesting how online activism can shape people's actions in the real world (Mourão and Brown 2022).

On the other hand, our findings suggest that offline confrontations can potentially shape subsequent online activism (Kim et al. 2017). Specifically, offline confrontational conflicts are significantly associated with more online conso-

nance and legitimacy with the protesters. This is consistent with the literature (Emmer et al. 2012): the success of real-world social movements such as #BlackLivesMatter, #Metoo, and Arab Spring, for instance, can often be attributed to the capacity to generate large-scale mobilization across different sectors of populations into actual collective action in various periods and locations. Other research has shown that one of the motivating factors of engagement in "non-normative" (more disruptive and potentially confrontational) action is the lack of efficacy of peaceful counterparts (Thomas and Louis 2014). While others found evidence for the "spillover effect," where offline participation spills over to online activism (Kim et al. 2017). The use of offline confrontational means can sometimes be more effective in generating awareness and support for social movements. Such tactics, while riskier, may capture public attention and support. For instance, protesters inciting violence, or damaging properties may increase the likelihood of social media sharing and reporting due to the disruptive nature of these activities. Those who see such activities as illegitimate may be less likely to show support for the protesters because they do not condone such disruption of order. Conversely, individuals who see government repressing peaceful protesters as unlawful or unjust are likely to share on social media to garner support and sympathy for the protesters (Mourão and Brown 2022). Our study illustrates the bidirectional relationship between online activism and offline confrontations is nuanced and multifaceted. Therefore, the effectiveness of social movements is not just the ideology mobilization (or when and where it happened), but also of the tactics employed and how these are perceived and amplified in the digital realm.

Recent work points out how social identity (or being in the minority group) is essential in studying social movements (Peay and Camarillo 2021). We, therefore, consider factors such as population, party affiliation, and racial composition. We find evidence from the covariance in our results, such as party affiliation and racial composition in the cities having BLM protests. Having a Democrat governor/mayor makes offline confrontations less likely. The higher the percentage of the White population in the given city, the less likely we will see offline confrontational events whereas a higher percentage of the Black population also decreases the likelihood of offline confrontational events. This could be interpreted as that, regardless of political leaning, individuals are less likely to join or show support for offline activities that are disruptive in nature. We do not see, however, a significant relationship between the percentage of the White or Black population with offline property damage. Lastly, we develop an annotation scheme and an online resonance ground-truth dataset. We present the first large-scale study of online resonance based on a few-shot machine learning model. Our research tool and dataset (subject to the terms of the data sources) will be made publicly available for future research.

Our study sheds light on the offline confrontations that tend to be overlooked as most literature primarily focuses on peaceful protests. The results also show social media's complex roles, especially in impacting the success or failure of movement organizational efforts and information or misinformation exposure. This highlights the importance of considering a broader range of factors, including the online public resonance, violence use in protests, and social identity, in understanding the dynamics of social movements. The inclusion of online resonance in our analysis provides a novel perspective on how digital platforms can influence and reflect real-world events, offering valuable insights for activists, policymakers, and researchers alike. By making our tools and datasets available, we aim to foster further exploration and understanding in this critical area of study.

Limitations and Future Work

Context Specificity. We derived our findings from events surrounding George Floyd's death and the subsequent BLM movements. The generalizability of our results to other contexts, like the 2021 Storming of the U.S. Capitol or the 2010 Arab Spring, remains uncertain. Although these events also witnessed significant confrontations, the nature and dynamics might differ. Future research can employ our methodology to evaluate the resonance-confrontation relationship in diverse scenarios, taking into account any confounding factors that may come into play.

Modeling Limitations. The model achieved an F1 score of 0.77 in few-shot label inference. However, the precision for two of the sparse labels, *Consonance-P* and *Dissonance-L*, was relatively low. This can potentially lead to analysis biases towards these two labels. To improve the accuracy of the label augmentation, a dataset that is richer in these labels would be beneficial. Also, our 19-day analysis relied on ARDL and SEM time series models. While other time series models exist, they often demand larger datasets for precise results. Comprehensive time-series data, spanning longer periods, would enhance our understanding of the underlying dynamics.

Data Constraints. (1) Offline Outcome: Given that the Crowd-Counting-Consortium dataset is volunteer-driven and mandates public verification, smaller confrontations might be overlooked. Besides confrontational events and property damage, alternative metrics, like the ratio of injuries to participants or confrontational to non-confrontational protests, can offer richer insights into offline confrontations. When more and richer data becomes available, exploring varied offline confrontation forms and their online activity interplay is crucial. (2) Online Outcome: Our online results stem from Facebook, a platform with its demographic and partisan biases (Diaz et al. 2016). Although our resonance measurement is adaptable, the results on Facebook might differ from other platforms. A promising avenue for future studies is analyzing the online-offline interplay across different platforms and discerning how such platform dynamics interact with offline events.

Potential Data Missingness. Our strategy for associating online resonance with locations relies on explicit location mentions in posts or metadata. Consequently, posts without clear location references might have been missed. Assuming that the rates of such omissions are consistent across time and place, the relative spatiotemporal outcomes might not be heavily impacted. Nonetheless, this assumption warrants validation in future studies.

8 Ethical Consideration

This research uses two data sources: the CCC and Facebook datasets. Data collected by the CCC is publicly available at the city/town level without any personal identifying information that has not already been reported in the public domain[4]. For the Facebook data, we follow CrowdTangle's Terms of Service[5] and use its official APIs to access 1) public contents in Groups and Pages and 2) post metrics such as "Like" count that the users posted or shared publicly and without restricting the audience.

As outlined in Section 7, our study has some limitations related to data representativeness and completeness. These limitations may skew the results of the phenomenon studied. The results should therefore be interpreted with caution.

This research sheds new light on the relationship between online activism and offline confrontations. However, the results of our study should not be interpreted as a simple correlation or even a causal link between online activism and any form of violent acts such as rioting. The line between protesting (a constitutionally protected form of expression) and rioting (a criminal act) is admittedly fuzzy and inconsistent. Sometimes the interpretation of a riot is influenced by the media, public discourse, implicit bias held by certain communities, and sometimes by the police's overreaction or excessive use of force against protestors (Simmons 2017; Anderson et al. 2022). It is also a formidable challenge to parse the motives of different protestors, rioters, or looters. Our study was unable to distinguish whether a confrontation truly interfered with legitimate law enforcement operations, and thus we use the term "confrontation" to indicate that both sides may initiate illegitimate acts. Our interpretation of the study results took a stance that a confrontational event may be related to the police's de-escalation strategy due to the overwhelming reports associated with the context of this particular movement (Amnesty International 2020), and it is crucial that confrontational/violent protests are interpreted with due nuance and context.

Acknowledgements

The authors would like to acknowledge support from AFOSR, ONR, Minerva, NSF #2318461, Collaboratory Against Hate Research and Action Center, and Pitt Cyber Institute's PCAG awards. The research was also partially supported by Pitt's CRC resources (RRID:SCR 022735 through NIH #S10OD028483). Any opinions, findings, and conclusions or recommendations expressed in this material do not necessarily reflect the views of the funding sources.

References

Amnesty International. 2020. USA: Law enforcement violated Black Lives Matter protesterśhuman rights, documents acts of police violence and excessive Force.

Anderson, J.; et al. 2022. Police Violence Against Black Protesters: A Public Health Issue. *Int'l J. Soc. Sci. Stud.*, 10.

Auxier, B.; and Anderson, M. 2021. Social Media Use in 2021. https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/.

Boulianne, S. 2018. Mini-publics and public opinion: Two survey-based experiments. *Political Studies*, 66(1).

Boulianne, S.; et al. 2020. Mobilizing media: Comparing TV and social media effects on protest mobilization. *Info., Comm, & Soc.*, 23(5).

Brady, W. J.; et al. 2021. How social learning amplifies moral outrage expression in online social networks. *Sci. Adv.*, 7(33): eabe5641.

Breuer, A.; et al. 2015. Social media and protest mobilization: Evidence from the Tunisian revolution. *Democratization*, 22(4): 764–792.

Brown, D. K.; and Mourão, R. R. 2021. Protest coverage matters: How media framing and visual communication affects support for Black civil rights protests. *Mass Comm. and Soc.*, 24(4): 576–596.

Cabrera, N. L.; et al. 2017. Activism or slacktivism? The potential and pitfalls of social media in contemporary student activism. *J. of Diversity in Higher Edu.*, 10(4): 400.

Chenoweth, E.; and Cunningham, K. G. 2013. Understanding nonviolent resistance: An introduction. *J. of Peace Research*, 50(3): 271–276.

Chiang, A. Y. 2021. Violence and the conditional effect of repression on subsequent dissident mobilization. *Conflict Mnmt. and Peace Sci.*, 38(6): 627–653.

Darwish, K.; et al. 2020. Unsupervised user stance detection on twitter. In *ICWSM Proc.*, volume 14, 141–152.

De Choudhury, M.; et al. 2016. Discovering shifts to suicidal ideation from mental health content in social media. In *CHI Proc.*, 2098–2110.

Devlin, J.; et al. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv:1810.04805*.

Dey, K.; et al. 2017. Twitter stance detection—A subjectivity and sentiment polarity inspired two-phase approach. In *ICDMW Proc.*, 365–372. IEEE.

Diaz, F.; et al. 2016. Online and social media data as an imperfect continuous panel survey. *PloS one*, 11(1): e0145406.

Drury, J.; and Reicher, S. 2005. Explaining enduring empowerment: A comparative study of collective action and psychological outcomes. *EU J. of Soc. Psy.*, 35(1): 35–58.

Edwards, P.; and Arnon, D. 2021. Violence on many sides: Framing effects on protest and support for repression. *British J. of Political Sci.*, 51(2): 488–506.

Elghazaly, T. 2016. Political sentiment analysis using twitter data. In *Proc. CCIOT*, 1–5.

Emmer, M.; et al. 2012. Changing political communication in Germany: findings from a longitudinal study on the influence of the Internet on political information, discussion and the participation of citizens. *Comm.*, 37(3): 233–252.

Ertugrul, A. M.; et al. 2019. Activism via attention: interpretable spatiotemporal learning to forecast protest activities. *EPJ Data Science*, 8(1): 5.

Greijdanus, H.; et al. 2020. The psychology of online activism and social movements: Relations between online and offline collective action. *Current Op. in Psy.*, 35: 49–54.

Gwet, K. 2001. Handbook of inter-rater reliability.

[4]https://sites.google.com/view/crowdcountingconsortium/submit-a-record

[5]https://www.crowdtangle.com/terms

STATAXIS Pub., 223–246.
Horne, G. M. 2022. *Protesting the Police: How Situational Threats Elicit Police Repression at Protest Events Targeting the Police*. East Carolina University.
Horowitz, J. 2022. Support for Black Lives Matter declined after George Floyd protests — Pew Research Center.
Horowitz, J.; et al. 2023. Views on the Black Lives Matter movement — Pew Research Center.
Howes, D.; and Classen, C. 2013. *Ways of sensing: Understanding the senses in society*. Routledge.
Hox, J. J.; and Bechger, T. M. 1998. An introduction to structural equation modeling.
Huet-Vaughn, E. 2013. Quiet riot: The causal effect of protest violence. *Available at SSRN 2331520*.
Joseph, K.; et al. 2021. (Mis) alignment between stance expressed in social media data and public opinion surveys. *arXiv:2109.01762*.
Kim, Y.; et al. 2017. The longitudinal relation between online and offline political participation among youth at two different developmental stages. *New Media & Society*, 19(6): 899–917.
Koopmans, R.; and Olzak, S. 2004. Discursive opportunities and the evolution of right-wing violence in Germany. *American journal of Sociology*, 110(1): 198–230.
Küçük, D.; and Can, F. 2020. Stance detection: A survey. *ACM CSUR*, 53(1): 1–37.
Liu, Y.; et al. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv:1907.11692*.
Maloney, E. K.; White, A. J.; Samuel, L.; Boehm, M.; and Bleakley, A. 2023. COVID-19 coverage from six network and cable news sources in the United States: Representation of misinformation, correction, and portrayals of severity. *Public Understanding of Science*, 09636625231179588.
Mazumder, S. 2018. The persistent effect of US civil rights protests on political attitudes. *American J. of Political Sci.*, 62(4): 922–935.
McGarty, C.; et al. 2014. New Technologies, New Identities, and the Growth of Mass Opposition in the Arab Spring. *Political Psy.*, 35(6): 725–740.
Mourão, R. R.; and Brown, D. K. 2022. Black Lives Matter coverage: How protest news frames and attitudinal change affect social media engagement. *Digital Journalism*, 10(4).
Muñoz, J.; and Anduiza, E. 2019. 'If a fight starts, watch the crowd': The effect of violence on popular support for social movements. *J. of Peace Research*, 56(4): 485–498.
Peay, P. C.; and Camarillo, T. 2021. No justice! Black protests? No peace: The racial nature of threat evaluations of nonviolent# BlackLivesMatter protests. *Soc. Sci. Quarterly*, 102(1): 198–208.
Ramos, A. M.; and Macau, E. E. 2017. Minimum sample size for reliable causal inference using transfer entropy. *Entropy*, 19(4): 150.
Schick, T.; and Schütze, H. 2020. Exploiting cloze questions for few shot text classification and natural language inference. *arXiv:2001.07676*.
Shojaie, A.; and Fox, E. B. 2022. Granger causality: A review and recent advances. *Annual Review of Stats. and App.*, 9: 289–319.
Shuman, E.; et al. 2022. Protest movements involving limited violence can sometimes be effective: Evidence from the 2020 BlackLivesMatter protests. *Proc. Nat. Acad. of Sci.*, 119(14).
Simmons, D. J. 2017. Patriots or Criminals?: An Experiment on How Media Framing Shapes Public Perception of Social Movements.
Spears, L. C. 2010. Character and servant leadership: Ten characteristics of effective, caring leaders. *J. of virtues & leadership*, 1(1): 25–30.
Steinert-Threlkeld, Z. C.; et al. 2022. How state and protester violence affect protest dynamics. *The J. of Politics*, 84(2): 798–813.
Sturmer, S.; and Simon, B. 2004. Collective action: Towards a dual-pathway model. *EU review of Soc. Psy.*, 15(1): 59–99.
Tajfel, H.; et al. 1979. An integrative theory of intergroup conflict. *Org. Id. : A reader*, 56(65): 9780203505984–16.
Thomas, E. F.; and Louis, W. R. 2014. When will collective action be effective? Violent and non-violent protests differentially influence perceptions of legitimacy and efficacy among sympathizers. *Personality and Soc. Psy. Bulletin*, 40(2): 263–276.
Thomas, E. F.; et al. 2014. Social interaction and psychological pathways to political engagement and extremism. *European J. of Social Psychology*, 44(1): 15–22.
VanderWeele, T. J.; and Ding, P. 2017. Sensitivity analysis in observational research: introducing the E-value. *Annals of Internal Med.*, 167(4): 268–274.
Wasow, O. 2020. Agenda seeding: How 1960s black protests moved public opinion and voting. *Amer. Political Sci. Review*, 114(3): 638–659.
Wasow, T. 2017. Generative grammar: rule systems for describing sentence structure. *The handbook of linguistics*.
Wei, K.; et al. 2020. Examining protest as an intervention to reduce online prejudice: A case study of prejudice against immigrants. In *Proc. WebConf*, 2443–2454.
Wilkins, D. J.; et al. 2019. Whose tweets? The rhetorical functions of social media use in developing the Black Lives Matter movement. *British J. of Soc. Psy.*, 58(4).
Williamson, V.; et al. 2018. Black lives matter: Evidence that police-caused deaths predict protest activity. *Perspectives on Politics*, 16(2): 400–415.
Yan, M.; Lin, Y.-R.; and Chung, W.-T. 2022. Are mutated misinformation more contagious? a case study of covid-19 misinformation on twitter. In *Proc. WebSci*, 336–347.
Yan, M.; et al. 2017. Quantifying content polarization on twitter. In *Proc. IEEE CIC*, 299–308. IEEE.
Yan, M.; et al. 2020. MimicProp: Learning to incorporate lexicon knowledge into distributed word representation for social media analysis. In *Proc. ICWSM*, volume 14.
York, C. B. 2022. How Media Influences the Use of Violence in Protests: An Analysis of the Black Lives Matter and# StopTheSteal Movements. *Inquiries J.*, 14(06).

9 Paper Checklist

1. For most authors...
   (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes.
   (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes. See Section 1, 3, 7, and 8.
   (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes. See Section 4 and 7.
   (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes. See Section 7 & 8.
   (e) Did you describe the limitations of your work? Yes. See Section 7.
   (f) Did you discuss any potential negative societal impacts of your work? Yes. See Section 7 and 8.
   (g) Did you discuss any potential misuse of your work? Yes. We discuss them in Section 8.
   (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes. We will release the dataset and model artifacts in the future.
   (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes.
2. Additionally, if your study involves hypotheses testing...
   (a) Did you clearly state the assumptions underlying all theoretical results? Yes. See Section 5.
   (b) Have you provided justifications for all theoretical results? Yes. See Section 6 - Sensitivity analysis and Robustness checks.
   (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? Yes. We tested the theories using structural equation modelings.
   (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? Yes. We conduct sensitivity analysis and robustness check to consolidate our findings. See Section 6.
   (e) Did you address potential biases or limitations in your theoretical framework? Yes. See Section 7 - Limitations
   (f) Have you related your theoretical results to the existing literature in social science? Yes. We discuss our theoretical foundations in Section 3.
   (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? Yes. We discuss the implications in Section 7.
3. Additionally, if you are including theoretical proofs...
   (a) Did you state the full set of assumptions of all theoretical results? N/A.
   (b) Did you include complete proofs of all theoretical results? N/A.
4. Additionally, if you ran machine learning experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? We will release them later for the anonymous review.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? Yes.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? Yes. See Figure 4.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? Yes. See Footnote 4.
   (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? Yes.
   (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? Yes. See Section 7 and 8.
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
   (a) If your work uses existing assets, did you cite the creators? Yes. See Section 4.
   (b) Did you mention the license of the assets? No. The software license is Apache 2.
   (c) Did you include any new assets in the supplemental material or as a URL? No.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? Yes. See Section 8.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes. We conform the social media policies. See Section 8.
   (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR? No. We need to conform social media platform policies sharing data.We can only share the public post IDs, without the data content itself.
   (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset? We will release the data without breaching the platforms' policies.
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
   (a) Did you include the full text of instructions given to participants and screenshots? N/A.
   (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? N/A.
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? N/A.
   (d) Did you discuss how data is stored, shared, and deidentified? N/A.

A Appendix

A.1 Study Pipeline and Acronyms

Our study consists of five major steps: **a)** collect the social media and protest data, and identify the geographical tags from their metadata; **b)** annotate ground-truth labels of posts' attitude toward protesters and police; **c)** use state-of-the art few-shots NLP model to augment the attitude labels for all posts; d) measure *consonance* and *dissonance* from the attitude labels by city by day, and report the trends in the U.S. cities with the most posts during the movement; **e)** employ time-series analysis to answer the proposed research questions.

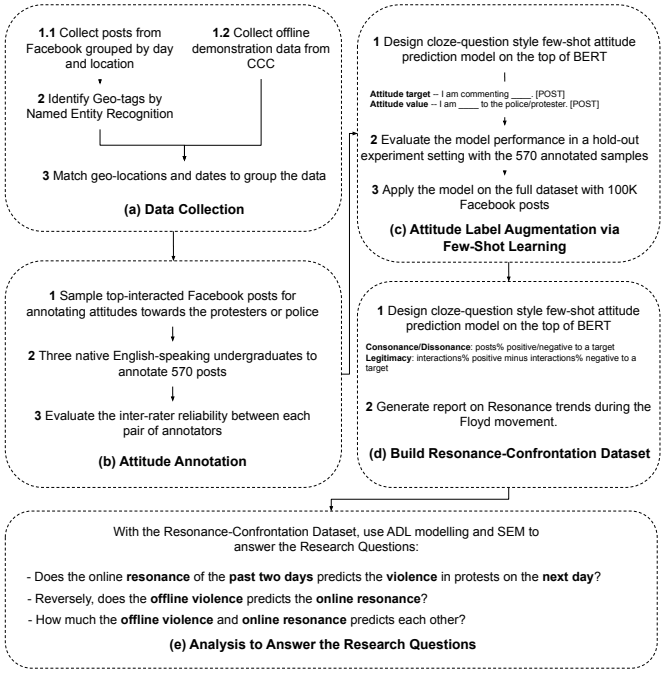

Figure 6: Flow Chart of the Study Pipeline.

The acronyms we employed in the study for resonance and legitimacy variables are listed in the following table:

Acronym	Description
Consonance-P	Consonance to **P**rotester
Dissonance-P	Dissonance to **P**rotester
Consonance-L	Consonance to **L**aw Enforcer (Police)
Dissonance-L	Dissonance to **L**aw Enforcer (Police)
Legitimacy-P	Legitimacy to **P**rotester
Legitimacy-L	Legitimacy to **L**aw Enforcer (Police)

Table 4: The acronyms for resonance and legitimacy.

A.2 Full Few-shot Learning Result Table

We report the full results of the few-shot attitude prediction in the following table:

	Fine-Grained Label		Ours		BERT Fine-tuning		N-gram		Coarse-Grained Label		Ours		BERT Fine-tuning		N-gram	
	Label	Ratio	F1	Aver. F1	F1	Aver. F1	F1	Aver. F1	Merged Label	Ratio	F1	Aver. F1	F1	Aver. F1	F1	Aver. F1
High-Resource	Favor Protester	13.6%	0.86		0.82		0.14		Favor-Protester / Against-Police	17.0%	0.81		0.70		0.43	
	Against Police	3.4%	0.00		0.00		0.00									
	Favor Police	0.6%	1.00		0.00		0.00		Favor-Police / Against-Protester	8.3%	0.36		0.33		0.33	
	Against Protetser	7.7%	0.22	0.52	0.23	0.50	0.33	0.34				0.77		0.74		0.65
	Neutral Protester	11.7%	0.49		0.42		0.30		Neutral	74.6%	0.81		0.80		0.73	
	Neutral Police	2.7%	0.22		0.00		0.34									
	No Target	60.2%	0.53		0.54		0.42									
	Label	Ratio	F1	Aver. F1	F1	Aver. F1	F1	Aver. F1	Merged Label	Ratio	F1	Aver. F1	F1	Aver. F1	F1	Aver. F1
Mid-Resource	Favor Protester	13.6%	0.64		0.52		0.29		Favor-Protester / Against-Police	17.0%	0.63		0.45		0.20	
	Against Police	3.4%	0.00		0.00		0.00									
	Favor Police	0.6%	0.67		0.00		0.00		Favor-Police / Against-Protester	8.3%	0.11		0.05		0.11	
	Against Protetser	7.7%	0.00	0.33	0.13	0.31	0.12	0.29				0.68		0.60		0.55
	Neutral Protester	11.7%	0.46		0.32		0.28		Neutral	74.6%	0.75		0.70		0.68	
	Neutral Police	2.7%	0.31		0.11		0.00									
	No Target	60.2%	0.30		0.32		0.34									
	Label	Ratio	F1	Aver. F1	F1	Aver. F1	F1	Aver. F1	Merged Label	Ratio	F1	Aver. F1	F1	Aver. F1	F1	Aver. F1
Low-Resource	Favor Protester	13.6%	0.19		0.08		0.00		Favor-Protester / Against-Police	17.0%	0.14		0.00		0.00	
	Against Police	3.4%	0.00		0.00		0.00									
	Favor Police	0.6%	1.00		1.00		0.00		Favor-Police / Against-Protester	8.3%	0.00		0.11		0.00	
	Against Protetser	7.7%	0.11	0.17	0.11	0.11	0.00	0.04				0.55		0.52		0.51
	Neutral Protester	11.7%	0.41		0.12		0.36		Neutral	74.6%	0.70		0.69		0.69	
	Neutral Police	2.7%	0.00		0.00		0.00									
	No Target	60.2%	0.14		0.12		0.00									

Table 5: Full Results of the Attitude Prediction. This table reports **a)** the detailed F1 scores per label and **b)** the weighted-average F1 scores for **i)** the N-gram baseline, **ii)** the BERT with fine-tuning, and **iii)** the few-shot learning models, in all training resource conditions. Overall the few-shot learning model (Ours as in the table) achieved the best performance, and reached $F1 = 0.77$ when trained with 273 samples.