Ertugrul et al. *EPJ Data Science* (2019) 8:5
https://doi.org/10.1140/epjds/s13688-019-0183-y

EPJ Data Science
a SpringerOpen Journal

EPJ.org

REGULAR ARTICLE

Open Access

Activism via attention: interpretable spatiotemporal learning to forecast protest activities

Ali Mert Ertugrul[1,2], Yu-Ru Lin[1*], Wen-Ting Chung[3], Muheng Yan[1] and Ang Li[1]

*Correspondence: yurulin@pitt.edu
[1]School of Computing and Information, University of Pittsburgh, Pittsburgh, USA
Full list of author information is available at the end of the article

Abstract

The diffusion of new information and communication technologies—social media in particular—has played a key role in social and political activism in recent decades. In this paper, we propose a theory-motivated, spatiotemporal learning approach, *ActAttn*, that leverages social movement theories and a deep learning framework to examine the relationship between protest events and their social and geographical contexts as reflected in social media discussions. To do so, we introduce a novel predictive framework that incorporates a new design of attentional networks, and which effectively learns the spatiotemporal structure of features. Our approach is not only capable of forecasting the occurrence of future protests, but also provides theory-relevant interpretations—it allows for interpreting what features, from which places, have significant contributions on the protest forecasting model, as well as how they make those contributions. Our experiment results from three movement events indicate that ActAttn achieves superior forecasting performance, with interesting comparisons across the three events that provide insights into these recent movements.

Keywords: Interpretable spatiotemporal learning; Event forecasting; Civil unrest; Protest activities

1 Introduction

Social movements are one of the most complex collective actions. They reflect how collectivities articulate and press a collectivity's interests to make significant changes in public policies and political decisions. Every day, news about social movement activity relevant to a variety of contested issues is being updated, on topics ranging from civil rights, to human rights, to gender equality, to gun control and others. Throughout human history, protests have been a primary means of engaging in social movements, in which collectivities usually give voice to their grievances and concerns about the rights and well-being of themselves and others [1]. In recent decades, the diffusion of new information and communication technologies—social media in particular—has reshaped the political activism of our time. From the Arab Spring, to the Occupy Wall Street movement, to the recent March for Our Lives gun violence protests, social media has been central in providing mobilizing information, coordinating demonstrations, and creating opportunities for people

to exchange opinions [2, 3]. In this work, our focus is whether and how online activities can forecast offline protests.

We started conceptualizing the prediction problem by considering what motivates people protest may help forecast; knowing the factors that drive people to protest may help to forecast demonstrations. Literature in social movements and social psychology has proposed theories and offered insights into *why* people protest [4–6]. For example, one fundamental factor of a given movement is its "connectedness," both in terms of how events connect with other events of a similar kind, temporally and spatially, and in terms of how they are embedded in an environment where people share similar sociocultural context. In other words, social movements are not merely instances of independent collective actions or protest events, but need to be investigated within their social, temporal and geographical contexts [1]. Empirically, however, in part due to the lack of proper analytical tools, studies (including social media studies) often analyze single events or movements via a case-study approach [7–10], or consider a large number of movement-related events independently of their relationships in time and space [11, 12].

It is crucial to move beyond single cases or aggregate measures and consider the dynamic interactions among the multitude of social, temporal and spatial dimensions. Analyses that are sensitive to spatial and temporal insertion will offer insights into *how* social movements were different in nature and in terms of progression. For example, some movements directly spoke to major national issues and garnered mass media coverage instantaneously, while others originated locally, relying on the efforts of ordinary advocates and grassroots activists before receiving media attention. To illustrate such differences, in this work we consider three recent movements—all of which connect to a similar social issue but are different in their progression in time and space. These include the Black Lives Matter (BLM) movement, which originated in the African-American community, and became nationally recognized during the protests and unrest in Ferguson, in August and November 2014 [13], as well as the marches that occurred following the white supremacist rally that took place in Charlottesville, in August 2017. The latter received intense media coverage immediately following the deadly attack that killed counter-protester Heather Heyer and President Trump's controversial statements [14]. As shown in Fig. 1, these different protest events left heterogeneous activity traces, both online and offline, over time and across locations, creating significant challenges in analyzing their spatial and temporal patterns.

Recent works in predictive modeling have shown considerable progress in predicting and forecasting spatiotemporal events, using machine learning methods such as transfer learning [15, 16]. However, most of them focus on prediction performance and lack the capability to facilitate understanding the nuanced spatiotemporal characteristics of social movement events. The theoretically-relevant questions include: in a movement, what social and activity features are associated with the subsequent events? To what extent are the local activities (observed from within a region) predictive of the subsequent events, compared to the global activities (observed outside of a region)? And what places' activities would have more far-reaching predictive power, in terms of signaling subsequent events in other places? None of the existing works have been able to answer these questions. In this work, we aim to provide a predictive modeling framework that is able to unveil the different spatiotemporal patterns and to answer these questions.

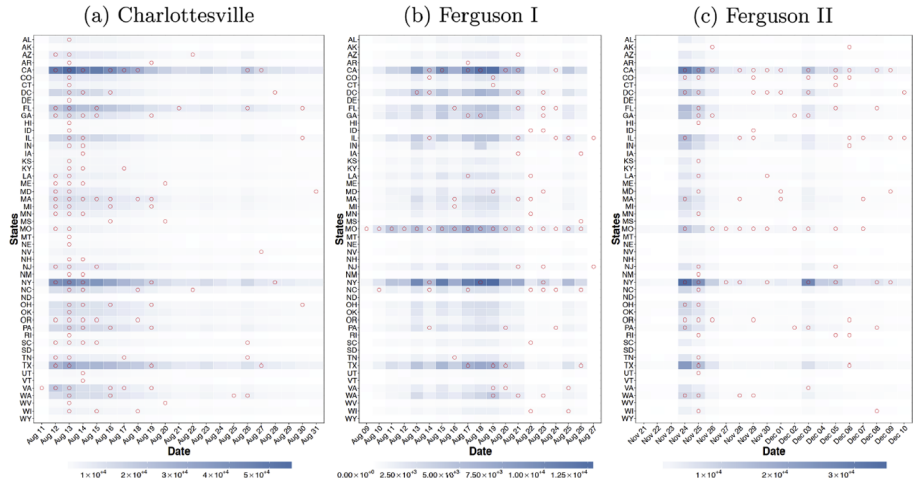


Figure 1 Spatiotemporal occurrence for different social movements by date (*x*-labels) and by location (*y*-labels). A red circle indicates at least one offline protest event happening on a particular day and state; the blue shade indicates the volume of tweets posted on the corresponding day/state. Charlottesville counterprotests exhibited burst patterns, in which most of the activities were sparked by a deadly violence attack and President Trump's statements on Aug 12th, 2017. In the first few days following the attack and the statements, more protest events occurred nation-wide and larger tweet volume was observed. The Ferguson I protests appeared to have a gradual build-up process, in which the activities were initially local (around Missouri and few states) following the shooting of Michael Brown on Aug 9th, 2014 and later received global attention. A global increase in tweet volume was observed until Aug 20th, 2014. The Ferguson II protests started on Nov 24th, 2014, with the announcement of the jury decision not to indict the police officer, and garnered global attention. The tweet volume for each state was greater in the first two days after the jury decision compared to the other days

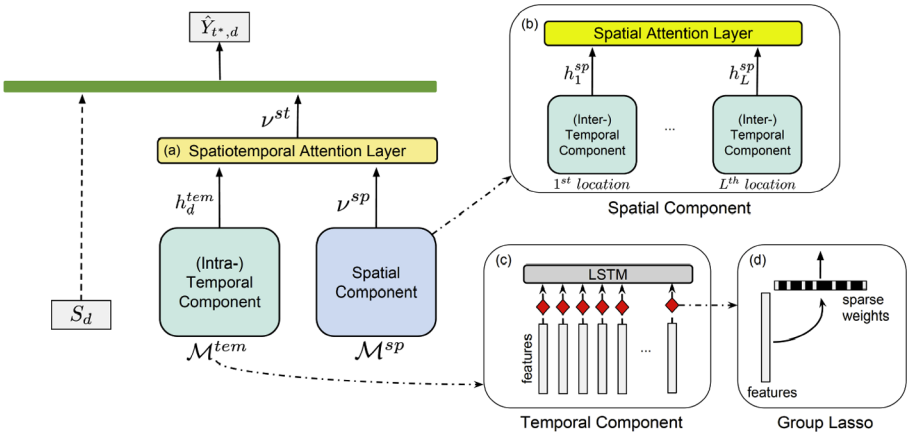


Figure 2 Overview of our proposed ActAttn architecture. It incorporates hierarchical attentional networks where the top level (**a**) differentiates the intra-region and inter-region importance, and the second level (**b**) identifies the hub regions. The temporal dependency of time-varying features in both intra- and inter-regions are modeled using LSTM (**c**), with sparse feature learning using Group Lasso (**d**)

Our proposed work. We propose a theory-motivated, spatiotemporal learning approach called *ActAttn* that addresses the aforementioned analysis challenge. Figure 2 gives an overview of ActAttn. Using social media and protest data, ActAttn seeks to characterize the social, spatial, and temporal features in relation to the subsequent protest activities in a unified and automatic manner. We develop a deep learning architecture that is not

only capable of forecasting the occurrence of future protests, but which also allows for interpreting what features, from which places, have significant contributions on the protest forecasting model, as well as how they make those contributions. To accomplish this, we introduce a two-level attentional network architecture that (a) differentiates the feature contribution from local (intra-region) and global (inter-region), and (b) identifies the regions, referred as the "hubs", that have a more salient contribution in predicting protest events globally. We utilize the lexicon approach to extract a range of linguistic features that allows for making sense of the association between the types of activity traces and future protests. We further leverage a sparse learning approach, Group Lasso [17], to select the compact set of features for enhancing the feature interpretability and generalizability.

Contributions. A major strength that differentiates our approach from the prior works is its *interpretability*. The interpretable capability comes from our model design, which has drawn largely upon prior social movement theories and empirical studies [1, 4–6]) regarding what motivates people to protest and what geological and sociocultural contexts and conditions may contribute to the inception and development of protests. The model design can be highlighted in terms of two aspects: (a) the selection of features, and (b) the differentiation of the predictive power that comes from local spatial patterns (or beyond).

To summarize, our contributions include: (1) *A unified, spatiotemporal leaning framework*: We propose a novel deep learning architecture, ActAttn, that automatically learns the relationship between the spatiotemporal activity traces observed from a broader community and the future protest events. This learning framework allows for principally comparing the spatiotemporal patterns from different movement events. (2) *Interpretability in hierarchical attention*: We use hierarchical attentional networks, together with Long Short-Term Memory (LSTM) [18], to model the temporal and spatial dependencies in the activity traces. The attentional networks allow for interpreting the importance of activities in different regions (intra- vs. inter-region contribution, and hubs), in terms of forecasting future events. This is the first model that differentiates the intra- and inter-region contributions in the spatiotemporal event forecasting domain. (3) *Interpretability in activity features*: We leverage Group Lasso to select a compact set of linguistic features, which allows for understanding the type of activity traces that are more reliably associated with future protests. (4) *Extensive experiments on forecasting performance, with in-depth analysis and comparison across three real-world movements*: We conduct extensive experiments on three social movement events: the counterprotests to the Charlottesville rally (August 2017), the first wave of Ferguson protests (August 2014), and the second wave of Ferguson protests (November 2014). Our results indicate a significant improvement in forecasting performance in comparison to several baseline and state-of-the-art methods. Moreover, we present in-depth analysis and comparison across three protest events in terms of their spatiotemporal characteristics and features. The results offer interesting insights regarding how social media "connectedness"—as operationalized at the level of features (social embeddedness) and the level of the model (the intra- vs. inter-region contribution)—could predict offline protest activity. Such analyses cannot be obtained with previous models. Finally, we have made our code and data available to ensure the reproducibility of our results.

2 Related work

2.1 Theoretical perspectives on antecedents of protest behaviors

Literature in social movements and social psychology offers us insights as to *why people protest*. Van Stekelenburgh and Klandermants [4, 5] proposed a motivational framework

that incorporates and synthesizes several sociopsychological factors that have been theorized and studied as critical to protests: (1) Identity: individuals' identification with certain groups/communities brings about a shared sense of future destiny and social responsibility; (2) grievance: a felt sense of illegitimate inequality; (3) emotion: emotions such as anger, guilt, fear, shame, and despair that "amplify" the felt grievance to be stronger and "accelerate" people to act more promptly; (4) social embeddedness: the social contexts one is exposed to and social networks one is embedded in—e.g., the more people engage themselves in the environment in which information about a certain grievance can be found, the more likely they are to start learning about the inequality and thus may take actions to protest or call for protests; and (5) efficacy: how one perceives that protests could make a difference.

In brief, protests are more likely to happen while people have the social interactions that offer more opportunities to learn about grievance and they emotionally resonate such illegitimate inequality, while, these people identify themselves as members of the communities that are affected by or responsible for the inequality, and while they believe protests could bring about change [4].

The framework aims to link the individual's psychological experiences—which are situated in certain types of social interactions, and which eventually lead to collective action and implications—and is particularly useful for our quantitative study. We are interested in Twitter users' individual tweeting behaviors, and whether the users are immersed in a kind of social embeddedness in which people who are seeking, sharing, and disseminating information about protests would come to gather together and linger. Such social embeddedness transforms individual grievance and emotion into their collective forms and may further facilitate the social actions of protests. We incorporate four factors—grievance, identity, social embeddedness, and emotion—into our model design and leverage the lexicon approach to operationalizing these factors (see details discussed in Sect. 3.3).

2.2 Forecasting protests and other events

There have been studies that employ social media data to examine social movements and unrest. Most of them followed a case-study approach in which descriptive statistics, regression analyses, or qualitative analysis were used for the exploration of movements [8, 9, 11, 19]. For example, Conover et al. [8] examined the temporal evolution of digital communication activity related to the Occupy Wall Street movement using Twitter-centric features including retweets, mentions, and user engagements. De Choudhury et al. [11] studied the temporal characteristic of social media participation and its relationships to offline protests related to BLM movement. Chung et al. [19] studied online social media discussions during the 2014 Ferguson protests, and employed a thematic analysis to differentiate tweets that engaged critical sensemaking from those solely focused on the event taking place. While these case studies provide detailed descriptions of the studied events, the analyses depend on specific questions of interests, and thus the results are sensitive to a particular data manipulation along the spatial or temporal dimensions.

There have been studies that utilize the spatial, temporal or spatiotemporal dependencies in modeling or predicting the events. Several studies employed logistic regression or heuristics to forecast/detect events from social media related to anomalies [20, 21], crime [22] and civil unrest [23, 24]. Cadena et al. [25] proposed an event forecasting model for civil unrest that uses a notion of activity cascades derived from the Twitter communication networks. Ning et al. [26] proposed a multiple instance learning based approach

that jointly forecasts protest events and identifies event precursors from news articles. Ramakrishnan et al. [27] proposed to forecast civil unrest from multiple data sources using models such as logistic regression with Lasso. Zhao et al. proposed spatiotemporal event forecasting through an enhanced Hidden Markov Model (HMM) [28] and multi-task learning [15, 16, 29]. Most of the existing techniques primarily focus on forecasting performance rather than interpreting spatiotemporal characteristics of social events. In addition, the potential interactions between temporal and spatial dimensions are often overlooked.

In terms of analyzing online social media content in the context of social movements, emotional commitment is the most widely studied factor. For example, De Choudhury et. al have used LIWC lexicon [30] to extract features that cover aspects of emotional expression, cognition, perception, social orientation, interpersonal awareness, and psychological distance [11]. On the other hand, the literature on why people protest (e.g., [4, 5]) has offered theoretical foundations and empirical evidence of what factors may be critical for protest occurrence and participation. In this work, we examine a set of new features that can provide theoretically-relevant interpretations about a social movement.

3 Method

3.1 Problem definition

Suppose there are L locations (e.g., cities, states) of interest, and each location l can be represented by a collection of static and dynamic features. The static features (e.g., population, political leaning) are features that remain the same or change slowly over a longer period of time, and the dynamic features (e.g., the percentage of tweets that express the "anger" emotion) are updated for each time interval t (e.g., hour, day). Let S_l be the set of static features of location l, and $X_{t,l}$ be the set of dynamic features for location l at time t. We are also given a binary variable $Y_{t^*,l} \in \{0, 1\}$ that indicates the occurrence of a future protest event for each location l at time t^*. The collection of dynamic features from all locations within an observing *time window* with size k up to time t can be represented as $\mathcal{X}_{t-k+1:t} = \{\mathcal{X}_{t-k+1}, \ldots, \mathcal{X}_t\}$, where $\mathcal{X}_{t'} = \{X_{t',1}, \ldots, X_{t',L}\}$.

Our goal is to predict the future event occurrence $Y_{t^*,l}$ at specific location l at a future time $t^* = t + \tau$, where τ is called the *lead time* for forecasting. The forecasting is based on the static and dynamic features of the location itself, as well as the dynamic features in the environment (from all other locations). Therefore, the forecasting problem can be formulated as learning a function $f(S_d, \mathcal{X}_{t-k+1:t}) \rightarrow Y_{t^*,d}$ that maps the input, the static and dynamic features, to a protest indicator at the future time t^* for a *target* location d.

To facilitate interpretation of the protest forecasting, we seek to develop a model that can differentiate the contribution of the features, the locality (local/intra-region features vs. global/inter-region features), and the overall importance of each location when contributing to the prediction of other locations. Therefore, we further organize the dynamic features $\mathcal{X}_{t-k+1:t}$ into two sets: the *intra-region* features, $\{X_{t-k+1,d}, \ldots, X_{t,d}\}$ represent the sequence of dynamic features for the location d, and the *inter-region* features, $\{X_{t-k+1,l}, \ldots, X_{t,l}\}$ for $l \in \{1, 2, \ldots, L\}$, contain the sequences of dynamic features for all locations of interest.

3.2 Model

As shown in Fig. 2, our proposed architecture involves three primary components: the temporal component $\mathcal{M}^{tem}$, the spatial component $\mathcal{M}^{sp}$, and the static features S_d. S_d pro-

vides location-specific information about the target location d. The temporal model $\mathcal{M}^{tem}$ is designed to model the contribution of the local dynamic features (*intra-region* features) for the target location. The spatiotemporal component $\mathcal{M}^{sp}$ is to model the spatiotemporal contribution of dynamic features for all locations of interest (*inter-region* features).

The recurrent unit. In both $\mathcal{M}^{tem}$ and $\mathcal{M}^{sp}$, we use LSTM as a building block in our model to capture the temporal relationships among the dynamic features. LSTM has been shown to be effective in capturing potential temporal dependency [31–33], and it addresses the vanishing and exploding gradient problems of basic recurrent neural networks (RNNs) by using explicit gating mechanisms (input, output and forget gates) to regulate the memory updates. We include a single LSTM network to model intra-region dynamics in $\mathcal{M}^{tem}$ (Fig. 2(c)). To capture the spatiotemporal relationship among all locations in $\mathcal{M}^{sp}$ (Fig. 2(b)), we include separate temporal components, each of which has the same structure as $\mathcal{M}^{tem}$. Each (inter-region) temporal component is then responsible for modeling the temporal dynamics of a single location. The LSTM outputs inside $\mathcal{M}^{tem}$ and $\mathcal{M}^{sp}$ are h_d^{tem} and $\{h_1^{sp}, h_2^{sp}, \ldots, h_L^{sp}\}$, respectively.

Hierarchical attention mechanism. An attention mechanism has been shown to be effective in reweighting the internal components in a neural architecture [34, 35]. We design a hierarchical attention mechanism to differentiate the importance of spatial and temporal information. First, in $\mathcal{M}^{sp}$, we incorporate a spatial attention layer on top of $\{h_1^{sp}, h_2^{sp}, \ldots, h_L^{sp}\}$ to learn the spatial importance among all locations (Fig. 2(b)). The idea is that not all the locations contribute equally to the prediction of event occurrence at a target location, and this attention layer is to reward the locations which contribute the most to correctly forecasting protest occurrence in the target location. The spatial attention is given by:

$$v^{sp} = \sum_l \alpha_l h_l^{sp}, \tag{1}$$

where v^{sp} is the spatial attention output that summarizes the aggregate contribution of all locations, and α_l is the attention weight for the location l to be learned based on a *Softmax* function. Second, we introduce a spatiotemporal attention layer to differentiate local (intra-region) and global (inter-region) feature contributions (Fig. 2(a)). The idea behind this layer is that, in some cases, the occurrence of protest events may largely depend on the temporal information within the locations themselves, while in other cases, the occurrence may depend more on the context of other locations or the global dynamics. The spatiotemporal attention layer is given by:

$$v^{st} = \alpha^{tem} h_d^{tem} + \alpha^{sp} v^{sp}, \tag{2}$$

where α^{tem} and α^{sp} are the attention weights corresponding to the outputs of temporal and spatial components, respectively. They are obtained at the output of the *Softmax* function. v^{st} is the spatiotemporal vector that aggregates the information learned from temporal and spatial dimensions. The forecasting of the occurrence of protest events is then given by:

$$\hat{Y}_{t^*,d} = \phi\left(W_c\left[S_d, v^{st}\right] + b_c\right), \tag{3}$$

where S_d is the static feature of the target location d, and W_c and b_c are the weight matrix and bias vector to be learned in the concatenation layer, respectively. ϕ is the activation

function where we apply the *Softmax* function in order to obtain posterior probabilities of occurrence and non-occurrence of the protest event.

Objective function. We incorporate the Group Lasso regularization into loss function. Group Lasso has been shown to be effective in several domains, such as robotic control [36] and multi-modal context [37] to select informative features. This regularization imposes sparsity on a group level, such that all the weights in a group are either simultaneously set to 0, or none of them are [17]. The main motivation for employing this regularization is to select informative features in temporal components (Fig. 2(d)) while assigning the optimal weights of the network at the same time. Therefore, it also enables us to interpret the model in such a way that redundant information from features are minimized, which allows for differentiating which features are important for the occurrence of protest events. The objective function is defined as:

$$\mathcal{L} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} Y_{ij}\log(p_{ij}) + \lambda_1 \left\| W^{tem} \right\|_{2,1} + \lambda_2 \sum_{l=1}^{L} \left\| W_l^{sp} \right\|_{2,1}, \tag{4}$$

where the first term is cross entropy loss, n is the number of samples, m is the number of class labels (event and non-event), and p_{ij} is the probability of the sample i being assigned to class j by the model. W^{tem} is the input weight matrix in $\mathcal{M}^{tem}$, and W_l^{sp} is the input weight matrix of (inter-region) temporal component of lth location in $\mathcal{M}^{sp}$. Note that the input weight matrix contains all weights of LSTM except for recurrent and bias weights. Moreover, λ_1 and λ_2 are the regularization factors for $\mathcal{M}^{tem}$ and $\mathcal{M}^{sp}$, respectively. Therefore, each component can be regularized by different factors. Group Lasso regularization can be written as:

$$\|W\|_{2,1} = \sum_{g\in G}\sqrt{|g|}\|g\|_2, \tag{5}$$

where g is the vector of outgoing connections (weights) from an input neuron, G denotes a set of input neurons, and $|g|$ indicates the dimension of g. We represent each input neuron in $\mathcal{M}^{tem}$ and in each (inter-region) temporal component of $\mathcal{M}^{sp}$ as a separate group so that G contains vectors of these groups.

3.3 Features

As mentioned earlier, there are two types of features: static and dynamic.

Static features reflect the political and demographic backgrounds of a location in which a protest event may take place, including the population of the state to which the location belongs (given as *population*), *population density*, *vote to Trump* (voting behaviors in 2016 presidential election as an indicator of the degree of conservativism in the location), and *region of the United States* (Northeast, Midwest, South and West). These features either remain unchanged or change slowly over time.

Dynamic features are to capture social media users' online activities that may be predictive of offline protests. Drawn upon social movement literature [4] (discussed in Sect. 2.1), we focus on four factors: *emotion*, *identity*, *grievance*, and *social embeddedness*.

Three dictionaries (LIWC [30], SentiSense [38], and Moral-Laden [39]) are used to capture the features indicating *emotions*, *grievance*, and *identity*, while additional relevant

features beyond these key factors are also included to test their usability. LIWC and SentiSense include a range of emotions, either positive or negative; LIWC offers the categories of *social* and *personal pronouns* that may serve as indicators of identity. The Moral-Laden dictionary is used with an attempt to capture grievance that results from the appraisal of relative deprivation based on moral rules; the dictionary is derived from moral foundation theory which suggests that humans engage in moral judgments along at least five dimensions: Harm/Care, Cheating/Fairness, Betrayal/Loyalty, Subversion/Authority, and Degradation/Purity. Some of the additional relevant features beyond these key factors discussed in literature are also included to test their usability.

Furthermore, in order to operationalize the type and level of *social embeddedness*, we caputre social media users' engagement in online discussion, including *number of tweets*, *number of reply tweets*, and *number of tweets with URL links*. Greater volumes of any of these tweeting behaviors (tweets, replies, and URLs) suggest that the public may be more aware of focal issues and events, and in turn be more motivated in seeking, spreading, and exchanging information, ideas, and emotions in cyberspaces. Such social contexts may raise individuals' perception of the efficacy of protests, which could lead to actual protest actions. More replies and URL links suggest being more embedded in relevant social networks. Replies suggest direct interactions with other embedded users. URL links, on the other hand, suggest information networks built based on relevant information/content created by others, including internal links with other tweets, and external links such as news, blogs, etc. The complete list of features and detailed interpretation are provided in Fig. 6(a), Fig. 6(b), and Sect. 5.2.

4 Experiments

4.1 Dataset

We choose social movements with social significance in order to test the design of our model with respect to the distinct social, temporal, and spatial dimensions of the nature of protests. Moreover, we choose movements in which the nature of the issues were relatively similar in order to compare and contrast the performance of the theory-driven features. Eventually, we select two movements: Black Lives Matter (BLM) and the counter-protests to Charlottesville's white supremacist rally. For BLM, we selected the two separate waves of protests regarding the police's killing of Michael Brown in Ferguson. The Ferguson unrests were symbolic protests under the umbrella of BLM in opposition to systemic racism against black people in the US. The Charlottesville counter-protests were the largest recent nationwide protest activities against white supremacism in the US.

Twitter data. We collected tweets with specific keywords or hashtags: the counter-protests to the Charlottesville rally [14], and the first and the second waves of the Ferguson protests [13]. The size and statistics of each dataset are provided in Table 1. *Charlottesville Dataset* was collected through the Streaming API based on 17 keywords and/or hashtags of interest.[a] Retweets were not included. These keywords were emerging during the event and were then widely used on Twitter to refer to the relevant issues and happenings. The *Ferguson I Dataset* and *Ferguson II Dataset* were collected based on the published work [40], using 45 keywords including #ferguson, #blacklivesmatter, "black lives matter" and the names of black people killed by police during 2014 and 2015. Based on the tweet IDs provided in the published dataset, we recollected the tweets within the two periods and excluded the retweets.

Table 1 Basic statistics of the datasets

Dataset	Duration	#Tweets	#Users	#Protest Occurrences
Charlottesville	Aug 11–Aug 31 (2017)	11.36M	5.93M	136
Ferguson I	Aug 9–Aug 27 (2014)	8.02M	2.76M	90
Ferguson II	Nov 21–Dec 10 (2014)	9.86M	3.80M	104

Protest data. We collected ground-truth data from the website of Elephrame[b,c] on the occurrence of offline protest events during the periods of the Charlottesville counter-protests and the two waves of the Ferguson protests. Elephrame provides information about civil unrest events which occurred in the US. This information is kept in a structured way and includes protest occurrence time (start date and end date), protest location (in state-level and city-level), protest subjects (sub-type of the protest event), description, number of participants, and at least one source link. We also incorporate news reports about BLM protests that were collected by the authors of [11]. Each piece of protest event information is based on the given source link(s). Note that there can be more than one event in the same location at the same time interval. In this work, we only consider whether an event occurred in a given location at that time interval, and we represent the occurrence using binary variables. As a result, we observed 136, 90 and 104 offline protest events during the three movements across the country.

Location extraction. In this work, we seek to forecast the occurrence of offline protest events at the state level, using Twitter users' activities. The locations of tweets are either extracted from their geocodes (if available) or inferred from the users' profiles. First, the geotagged tweets posted from the United States include state information in their 'place' field. These kinds of posts include either a state name or state code. We directly use this information as the location indicator. Second, we find the location information of the tweets from user profiles. We follow this approach for the tweets whose locations cannot be identified using the first approach. Similar to the first approach, we identify the locations (state name or state code) if they are explicitly written in the user profiles. If they are not, we also look for the names of cities located in the United States. If we identify a city name in the profile, we map it to its corresponding state. For this purpose, we use a dictionary including city-state pairs in the United States from Encyclopedia Britannica.[d] Note that there can be more than one city with the same name in different states. Therefore, we discard such cities in this study. In total, we were able to extract tweet locations at the state level for 29.9%, 41.5% and 43.3% of all tweets in the Charlottesville, Ferguson I, and Ferguson II datasets, respectively.

4.2 Comparison methods and settings

We compare our approach with several state-of-the-art approaches as the baseline methods. In order to evaluate the forecasting effectiveness of the proposed model, we select three sets of baseline methods.

The first set includes Logistic Regression (LR) and Support Vector Machine (SVM) classifiers, since they are widely-used machine learning methods in the event detection/forecasting literature. With these methods, we examine the effect of static, intra-region and inter-region features by combining all features together. The second set of methods include recently-developed neural-network-based models, such as RNNs and LSTMs in particular, as they have been shown to have superior performance in event

forecasting problems due to their capability of modeling the temporal dependencies. The third set of methods are the state-of-the-art spatiotemporal event forecasting approaches recently proposed by [15], including regularized multi-task feature learning *(RMTFL)*, constrained multi-task feature learning I *(CMTFL-1)* and constrained multi-task feature learning II *(CMTFL-2)*. These methods formulate event forecasting for multiple locations as a multi-task learning problem. They build event forecasting models for different locations simultaneously by restricting all locations to select a common set of features. Note that none of the existing approaches support the hierarchical structure of features coming from intra- and inter-regions, and we will discuss the importance of such differentiation more in Sect. 5. The baseline methods are summarized as follows:

The first set:

- *Logistic Regression (LR)* is simple LR model. We have three baselines for this model. $LR[tem]$ uses only intra-region features, $LR[s, tem]$ concatenates static and intra-region features, and $LR[s, tem, st]$ merges all features as the input.
- *Support Vector Machine (SVM)* is simple SVM model. $SVM[tem]$ employs only intra-region features, while $SVM[s, tem]$ combines static features with intra-region features. Also, all features are used as input in $SVM[s, tem, st]$.

The second set:

- *LSTM* is a basic LSTM network that employs only intra-region features. It does not consider static features and spatial relationships among regions.
- *S + LSTM* is the model where intra-region features are given as inputs to the LSTM network. Then, the embeddings of dynamic features is concatenated with the static features. This model does not consider the spatial relationships among regions.
- *S + LSTM (GL)* has the same structure as *S + LSTM*, yet it is trained incorporating Group Lasso regularization. With this model, we aim to monitor the effect of Group Lasso regularization on the performance of the *S + LSTM* model.

The third set:

- *RMTFL* employs a regularization parameter to control the model sparsity.
- *CMTFL-1* introduces a constraint to control the number of features in the model for sparsity.
- *CMTFL-2* restricts the number of features selected from static and dynamic groups separately.

Furthermore, to evaluate the effectiveness of individual components of ActAttn, including the Group Lasso regularization and hierarchical attention mechanism (spatial and spatiotemporal attentions), we include several variants of ActAttn for comparison as follows:

- *ActAttn (w/o GL)* has our proposed structure, yet Group Lasso regularization is not applied during training.
- *ActAttn (w/o stAttn)* does not include the spatiotemporal attention layer; instead, h_d^{tem} and v^{sp} are concatenated.
- *ActAttn (w/o spAttn)* does not include the spatial attention layer; instead, a linear projection layer is used.

Settings. In the experiments, we use 'day' as the time unit and 'state' as the location unit. The last five days from each dataset are used as the test sets, and rest as the training sets. The training set of the *Charlottesville* dataset contains 127 protest events (15.6% of all samples in the training set) and the test set contains 9 events. The training set of the *Ferguson I* dataset contains 63 protest events (9% of all samples in the training set) and the

test set contains 27 events. The training set of the *Ferguson II* dataset contains 82 protest events (10.7% of all samples in the training set) and the test set contains 22 events. We enumerate different settings of window size and lead time. The window size k is set to be $\{1, 2, 3\}$ and the lead time τ is set to be $\{1, 2, 3\}$. The hidden unit size for LSTM is 16. The architecture is trained using the Adam optimizer [41] with a learning rate of 0.001. For the models incorporating Group Lasso regularization, regularization factors λ_1 and λ_2 are selected from the set $\{10^{-5}, 10^{-4}\}$. During test time, the input weights with absolute values smaller than 10^{-3} are set to 0 as suggested in [17]. Our code and data are available at https://github.com/picsolab/actattn. For the state-of-the-art *MTFL*-based models, the regularization parameter is set to be $\{10^{-4}, 10^{-3}, \ldots, 10^{3}, 10^{4}\}$. The number of features to be selected in the *CMTFL*-1 model is set to be $\{5, 10, \ldots, 55\}$. The numbers of static and dynamic features to be selected in the *CMTFL*-2 model are set to be $\{4, 5, 6, 7, 8\}$ and $\{5, 10, \ldots, 50\}$, respectively.

5 Results

In this section, we present a comprehensive set of results. First, in Sect. 5.1, we show the forecasting effectiveness of the proposed model in comparison with the baseline and state-of-the-art forecasting approaches, and based on the aforementioned experiment settings. In Sect. 5.2, we analyze different kinds of predictive features identified by our model and interpret their effects in relation to the theoretical factors. In Sect. 5.3, we analyze and interpret different kinds of spatial contributions (intra- vs. inter-region). Finally, in Sect. 5.4, we explore the potential of using additional content features in the current forecasting framework.

5.1 Performance comparison

We compare the forecasting performance of ActAttn with the comparison methods. We organize the results to answer the following three questions:

1. Overall, how well could ActAttn forecast future protest event occurrences, compared with the baseline methods? (Sect. 5.1.1)
2. As missing information is common in social event predicting problems, how robust is ActAttn in dealing with missing information, compared with the baseline methods? Additionally, will ActAttn's spatiotemporal architecture help deal with the missing or noisy information? (Sect. 5.1.2)
3. How early in time can ActAttn effectively predict future protest event occurrences? (Sect. 5.1.3)

5.1.1 Overall performance

As shown in Table 2, the results indicate that ActAttn achieves the highest F-score and AUC values on the Charlottesville (0.400 and 0.843), Ferguson I (0.462 and 0.822) and Ferguson II (0.471 and 0.853) datasets. The F-scores for all methods are low due to the imbalance in class distribution (9%–15% protest events). Further, while the protest occurrence pattern is different for each dataset (Fig. 1), ActAttn is robust with respect to various distribution of the data, and is able to model temporal and spatial dimensions under various conditions successfully.

We show the significance of static features by comparing the results of $LR[tem]$ with $LR[s, tem]$, $SVM[tem]$ with $SVM[s, tem]$, and $LSTM$ with $S + LSTM$. It can be seen that,

Table 2 Forecasting results

	Charlottesville		Ferguson I		Ferguson II	
	F-score	AUC	F-score	AUC	F-score	AUC
LR[*tem*]	0.200	0.696	0.103	0.733	0.343	0.752
LR[*s*, *tem*]	0.182	0.789	0.259	0.766	0.327	0.789
LR[*s*, *tem*, *st*]	0.200	0.734	0.230	0.722	0.314	0.773
SVM[*tem*]	0.200	0.818	0.000	0.791	0.400	0.816
SVM[*s*, *tem*]	0.186	0.809	0.000	0.796	0.408	0.837
SVM[*s*, *tem*, *st*]	0.000	0.782	0.000	0.754	0.313	0.780
LSTM	0.240	0.752	0.415	0.801	0.417	0.819
S + *LSTM*	0.267	0.778	0.423	0.804	0.439	0.838
S + *LSTM* (*GL*)	0.308	0.793	0.423	0.805	0.440	0.839
RMTFL	0.182	0.663	0.250	0.703	0.250	0.829
CMTFL-1	0.182	0.664	0.350	0.711	0.316	0.805
CMTFL-2	0.200	0.661	0.333	0.711	0.324	0.815
ActAttn (*w/o GL*)	0.308	0.830	0.459	0.820	0.464	0.849
ActAttn (*w/o stAttn*)	0.324	0.797	0.406	0.783	0.409	0.842
ActAttn (*w/o spAttn*)	0.333	0.836	0.448	0.812	0.448	0.846
ActAttn	**0.400**	**0.843**	**0.462**	**0.822**	**0.471**	**0.853**

in nearly all cases, combining static features with intra-region features yields better F-score and AUC values. When we further combine inter-region features, we observe that *LR*[*s*, *tem*, *st*] and *SVM*[*s*, *tem*, *st*] give worse results compared to *LR*[*s*, *tem*] and *SVM*[*s*, *tem*], respectively. Thus, these models fail to capture the spatiotemporal information from the concatenated inter-region features. In our approach, combining inter-region features with static features and intra-region features increases the performance in all *ActAttn*-based methods except *ActAttn (w/o stAttn)*. Moreover, *S* + *LSTM* (*GL*) performs slightly better than *S* + *LSTM* and eliminates some of the redundant inputs in all three models.

To compare the performance of ActAttn with the state-of-the-art spatiotemporal event forecasting approaches, we performed experiments on all the datasets with *RMTFL*, *CMTFL*-1 and *CMTFL*-2 proposed by [15] by employing various parameter combinations. We report the best test performances of these approaches on each dataset. The results indicate that ActAttn significantly outperforms all three approaches on all datasets in terms of both F-score and AUC values.[e]

To examine the effect of Group Lasso regularization and the hierarchical attention mechanism, we compared the performance of *ActAttn* to its three variants. Although *ActAttn* slightly outperforms *ActAttn (w/o GL)*, Group Lasso regularization provides sparsity and selection of a compact set of features. The *ActAttn* model provides 95.0%, 76.6% and 96.8% sparsity for Charlottesville, Ferguson I and Ferguson II, respectively. It is computed as the ratio of zero input weights over the total number of input connections. Furthermore, we compare *ActAttn* to *ActAttn (w/o stAttn)* and *ActAttn (w/o spAttn)* to examine the effect of the hierarchical attention mechanism. We observe that *ActAttn* performs significantly better than *ActAttn (w/o stAttn)*. This shows the importance of the spatiotemporal attention layer which adjusts the local and global feature contributions. Similarly, *ActAttn* performs superior to *ActAttn (w/o spAttn)*. Removal of the spatial attention layer from the proposed architecture also results in loss of interpretation capability about the most contributing locations. Our results reflect that incorporating spatiotemporal attention layer enhances the performance of the model the most.

5.1.2 Robustness to missing information

A common challenge in predicting/forecasting social events is that data (including but not limited to social media data) often involve missing information or are only partially complete. For example, social media user activity may be sparse in a certain region or at a particular time. As ActAttn was designed to capture the spatiotemporal characteristics and features, we expect that ActAttn would be more robust to missing data if the model effectively captures the spatiotemporal structure from the training data. To test this, we simulate two kinds of missing information scenarios.

(1) *Missingness in time and space*: A missing value could occur in any feature of any region at any time. To simulate this, we randomly removed different levels of input data (20%, 40%, 60% and 80%) from the test sets. We then filled the missing values by randomly assigning values taken from the range of non-missing values of the corresponding features. In this setting, the comparison methods include those methods that take all features (static, temporal and spatial features) as input and have the best overall performance within each of the method variants. Figure 3 shows the forecasting performances of the methods for each dataset over different levels of missing data. The results indicate that

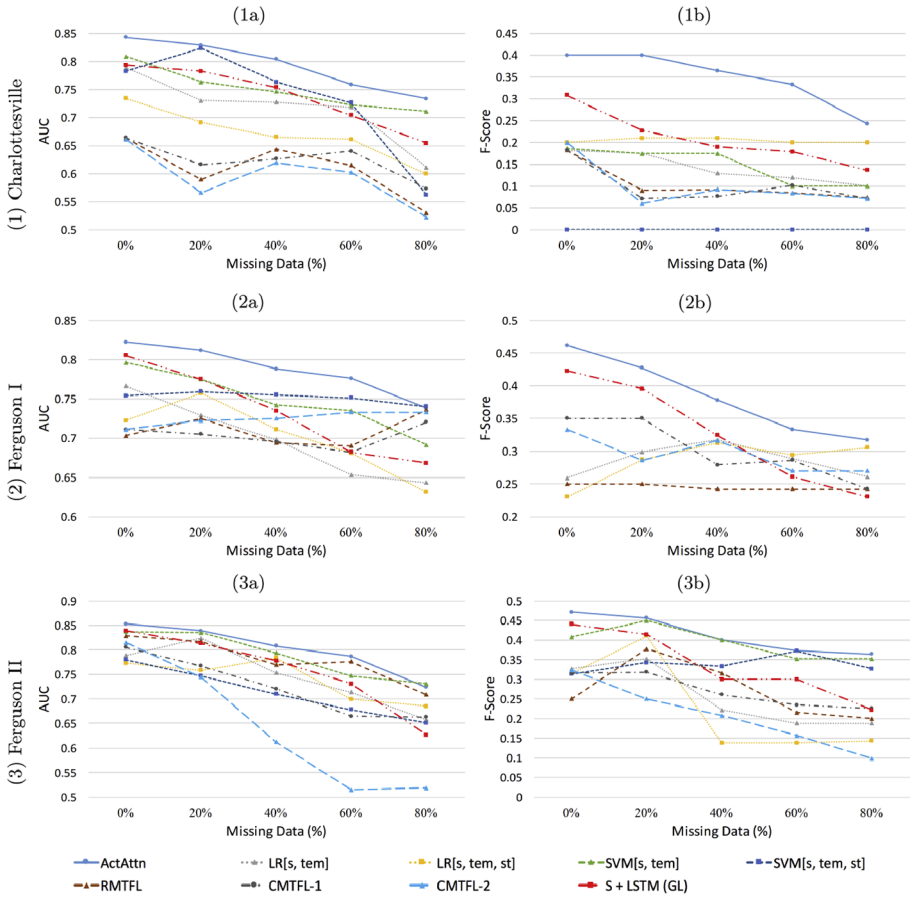


Figure 3 Forecasting results against varying levels of missingness (in time and space) from the test sets. The *x*-axes indicate the levels of missingness, and the *y*-axes indicate the performance in terms of (**a**) AUC and (**b**) F-score results

ActAttn performs significantly better (in terms of both AUC and F-score) than all the other methods on all datasets and for almost all levels of missing data.

(2) *Missingness in certain regions*: The missing values could occur in a particular region for an entire (short- or long-term) period of time. To simulate this, we randomly selected different proportions of regions (states, ranging from 20% to 80%) and removed their inputs entirely from the test sets. The removed regions thus do not contribute to forecasting events in any of the target regions. In this setting, we included the methods taking features from the other states for comparison. Note that although these methods include features from the other states, they do not differentiate intra- and inter-region contributions. Therefore, we expect that these comparison methods may suffer from missing some degree of regional input. Figure 4 shows the forecasting performance of the methods for each dataset over different levels of missing region information. The results show that ActAttn outperforms the other methods in terms of both AUC and F-score on all three datasets and for all levels of missing region information. Also, we observed that ActAttn performs more stable in nearly all conditions.

In both scenarios, we observe that ActAttn is more robust compared to other methods. This suggests that the design of ActAttn is particularly useful in dealing with missing information—the hierarchical attention mechanism learns important regions and sum-

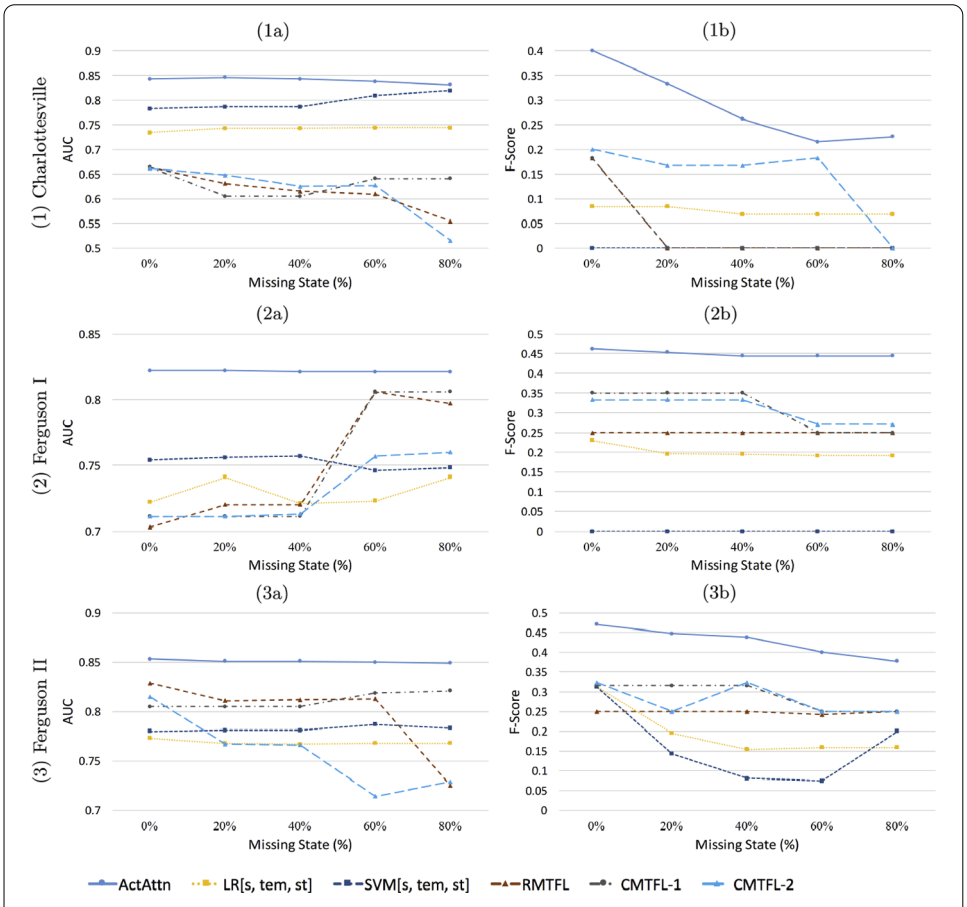


Figure 4 Forecasting results against varying levels of missingness for regions (states) from the test sets. The *x*-axes indicate the levels of missingness, and the *y*-axes indicate the performance in terms of (**a**) AUC and (**b**) F-score results

marizes the spatiotemporal information from intra-region and inter-region features, and the Group Lasso regularization imposes sparsity and selects an informative set of features.

5.1.3 Performance analysis with varying lead time

To examine how early in time ActAttn effectively forecasts future protest event occurrences, we tested the forecasting under different *lead time* conditions. A lead time τ is the length of time (number of days, in our experiment) from which the data are available for forecasting events occurring at $t + \tau$ (as defined in Sect. 3.1). We evaluated our method with different lead time settings, where $\tau \in \{1, 2, 3\}$. Figure 5 shows the forecasting performances of ActAttn and comparison methods over different lead time settings. The results indicate that ActAttn has significantly better performance compared to other methods in terms of AUC and F-score on three datasets across almost all lead time settings. This suggests that ActAttn is able to achieve better and more stable performance for short-term event forecasting, up to $\tau = 3$. Due to the limitation of our data, we do not examine longer-term event forecasting in this work.

We further examine the performance results for ActAttn with different window size k and lead time τ. As defined in Sect. 3.1, the window size represents the amount of information needed for forecasting in terms of the number of consecutive days as input. The AUC

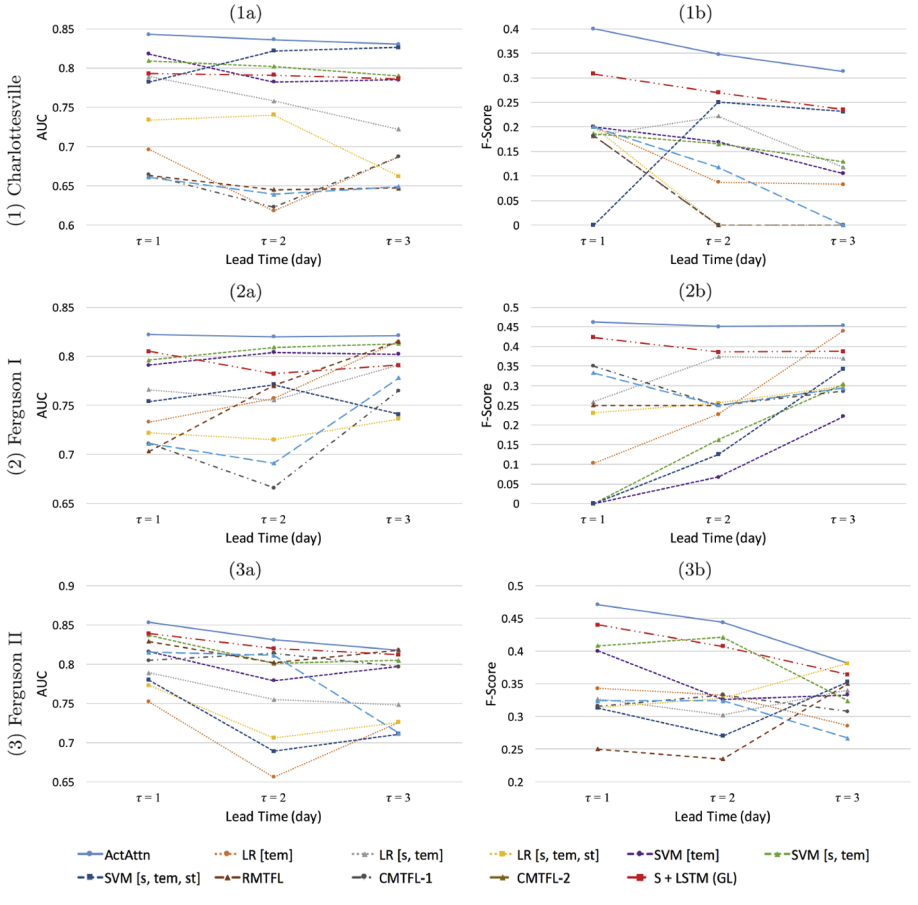


Figure 5 Forecasting results against different lead times. The *x*-axes indicate lead time τ, and the *y*-axes indicate the performance in terms of (**a**) AUC and (**b**) F-score results

Table 3 AUC results of ActAttn with respect to different window size k and lead time τ

	Charlottesville			Ferguson I			Ferguson II		
	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
$\tau=1$	0.842	**0.843**	0.823	0.807	0.815	**0.822**	**0.853**	0.832	0.800
$\tau=2$	0.839	0.836	0.823	0.807	0.820	0.820	0.831	0.836	0.832
$\tau=3$	0.830	0.830	0.819	0.791	0.808	0.821	0.818	0.820	0.811

values for corresponding results are given in Table 3. Accordingly, the best performances are achieved when ($k=2, \tau=1$), ($k=3, \tau=1$) and ($k=1, \tau=1$) for the Charlottesville, Ferguson I and Ferguson II models, respectively. In general, the performance either remains stable or decreases slightly with an increase in the lead time τ, regardless of window size k.

5.2 Interpreting the impact of features

We interpret the significance of features, organized by *intra-region*, *inter-region*, and *static*. Group Lasso regularization has selected a subset of features with the most discriminative power in the models.

5.2.1 Intra-region dynamic features

Which dynamic features of a state were most important for predicting future protests in the same state? Figure 6(a) gives a summary, and we provide our interpretation below. To better understand the significance of those features in each protest context, a manual inspection of the tweet content is conducted.

1. *Social Embeddedness*. Among the three relevant features (number of tweets, number of replies, and number of tweets with URLs), *num_tweets* is the most powerful that for all of the three protest events, online activism within a state is predictive of future offline protests in the same state. *Num_urlTweet*, which indicates the number of Twitter posts that contain an external link to other sources, is also found to be a useful predictor—except in the case of Ferguson I. This may be caused by the fact that Michael Brown's death was initially paid little attention by news outlets, so the external news or relevant URLs may be less indicative of online activist engagement.

2. *Emotions*. Both *positive* and *negative emotions* (*posemo* and *negemo* from LIWC), are important in all models. Particularly, *anger* (from LIWC) is predictive for all, which suggests that anger is a good indicator in predicting protest for all cases. Moreover, certain emotions stand out for each protest scenario. For example, *disgust* (from SentiSense) is predictive in Charllottesville; *hate* (from SentiSense) in Ferguson I; and *fear* (from SentiSense) in Ferguson II.

In addition, a Moral-Laden feature, *PurityVice* (the extent of impurity and corruption) unexpectedly captures an intensely *annoying* emotion in predicting Ferguson I protests. We uncovered this when analyzing the relevant tweets, in which the online community extensively express its sense of being "*sick of*" or feeling "*disgust*" for the fact that another black life was taken by the police.

3. *Grievance.* Our results indicate that Moral-Laden features are not able to capture grievance. However, through further analysis of the feature *negation* (from LIWC)—the use of words such as *no*, *not*, *never*—suggests it may serve as an indicator of grievance. This feature is important for all models, and especially for Ferguson I and II. Negation is used in online communities to emphasize appraisals of how unbelievable and unrealistic a

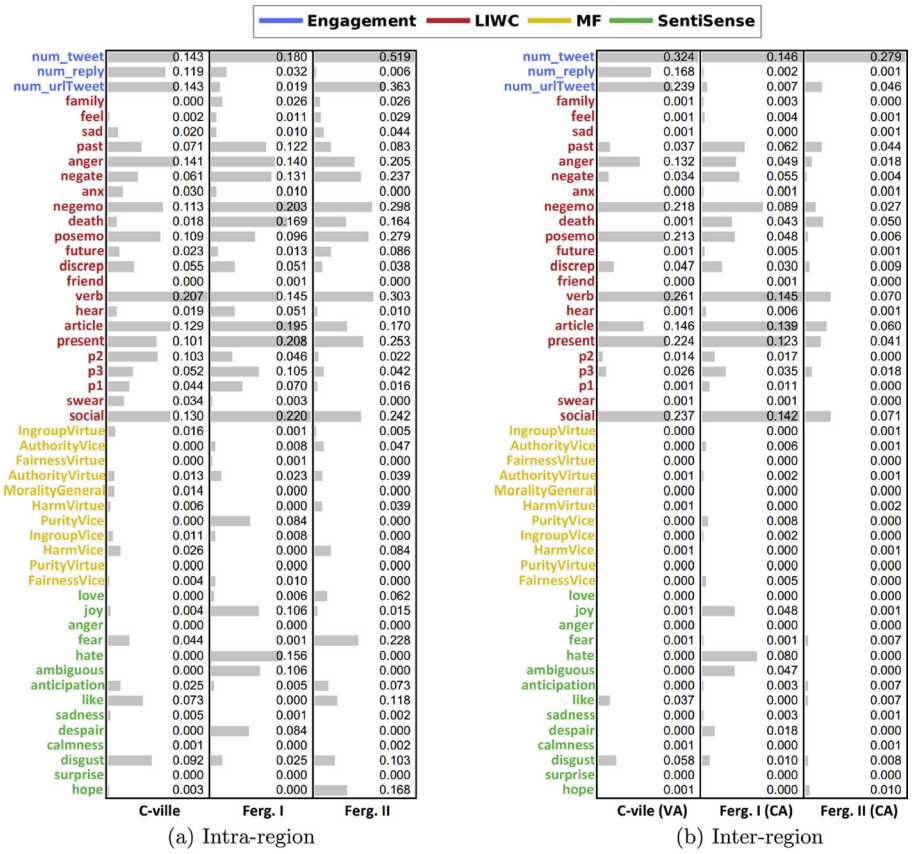


Figure 6 Mean absolute values of intra-region and inter-region input (gate) weights. These are the input weights learned from the neural network model (the LSTM networks in the temporal and spatial components) and the magnitude of weights (which can take any values) allows for a comparison of the relative importance of different features. (**a**) Intra-region input weights. (**b**) Inter-region input weights

situation is when they learn about the specific happenings (e.g., the shooting of unarmed Michael Brown, the grand jury's decision to not indict Officer Wilson, and a public rally against racism) that strongly conflict with their normal sense of moral principles, which indicates grievance (referring to the feeling of illegitimate injustice).

4. *Identity*. *Social* (from LIWC), which refers to the use of personal pronouns—especially plural ones such as *we*, *you*, *they*, and *people*—is predictive for all models. These terms are extensively used to call upon in-group members (we) to recognize the grievances and express protesting voices against out-group members (they; e.g., the police, a group considered by a majority of the online community as an embodiment of racism).

5. *Others*. We also observed the impact of other features. The features of both *verb* (from LIWC) and *present* (from LIWC) are important in all cases, which indicates the use of verbs (especially present tense of both auxiliary verbs, such as *is*, *are*, *have*, and *can*) to emphasize the happenings and perceived grievance as serious matters of fact. We also observed the use of action verbs such as *go*, *take*, *make*, *need*, and *think*, which call for necessary actions.

The features of *personal pronouns* (from LIWC) are also significant predictors, which involve the reference of and discussion of certain people at the center of why people protest for or against. For example, *you* is important for Charlottesville; the second-person pro-

noun extensively refers to President Trump, as online activists questioned him earnestly about his position on racism. Likewise, *he* is important in predicting Ferguson I protests, which is used to refer mostly to either Michael Brown or Eric Garner, both of whom were killed by the police; *they* refers primarily to the police. In Ferguson II, online activists focused more on the judicial system, which was seen as unsuccessful in delivering justice. Thus, personal pronouns are less predictive.

5.2.2 Inter-region dynamic features

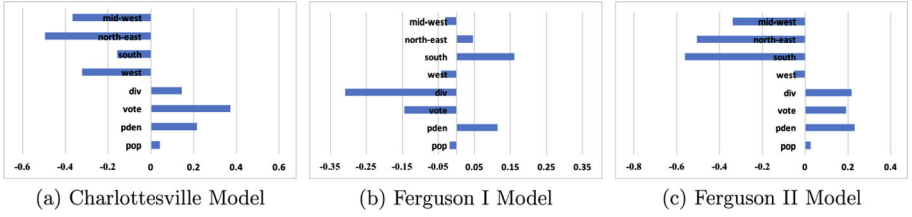
We explore the effectiveness of inter-region dynamic features by analyzing the input weights (only the portions which connect inputs to input gates) of each temporal component in spatial component, $\mathcal{M}^{sp}$. Figure 6(b) summarizes the importance of inter-region dynamic features in predicting protest within given states. Large percentages (96.5%, 77.6%, and 97.9% in the cases of Charlottesville, Ferguson I and Ferguson II, respectively) of the input weights are discarded as a result of Group Lasso regularization. We select Virginia (VA) from the Charlottesville, California (CA) from the Ferguson I and CA from the Ferguson II models, to analyze the inter-region input weights because these states are all 'hub' states for corresponding models (explained in Sect. 5.3). The result suggests that other states' features are much less predictive, especially for Charlottesville and Ferguson II. *num_tweet* performs exceptionally well, which indicates that online community activities in other states could be also significant across all other states.

5.2.3 Static features

Figure 7 shows the importance of static feature weights in the three models. The features representing US regions indicate how predictive the region class for a given state is—e.g., is a state in the South more or less likely to have future protests? The results of the Charlottesville and Ferguson II models exhibit similar patterns, suggesting that both protest events took place more all over the US, while Ferguson I started locally with a majority of black communities, and its model shows that being a Southern state itself is predictive of future protests.

5.3 Interpreting the local and global contributions and hubs

ActAttn enables us to explore the proportion of local (intra-region) and global (inter-region) contributions in forecasting protest events, and allows for discovering the "hubs" that have a more salient contribution in predicting protest events globally. The intra- and inter-region contributions can be identified based on the spatiotemporal attention weights



(a) Charlottesville Model (b) Ferguson I Model (c) Ferguson II Model

Figure 7 Values of static feature weights. These are the static feature weights learned from the neural network model. The weights (which can take any values) allow for a comparison of the relative importance of different features. (**a**) Charlottesville model. (**b**) Ferguson I model. (**c**) Ferguson II model

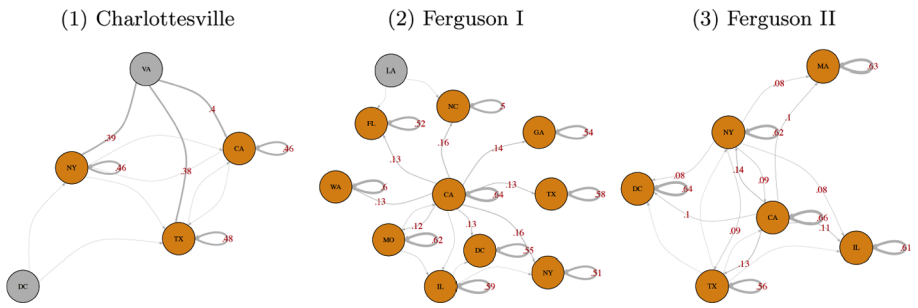


Figure 8 Exploration of local and global contributions to forecasting. While the orange nodes represent the states which are correctly predicted by the corresponding models, the gray nodes denote the states either not correctly predicted or where no events occurred, yet still contribute to forecasting events in the correctly predicted states with a value above a certain threshold. The edges indicate the contribution to forecasting from source state to target state. The thicker the edge, the more the contribution

in our model, and the hubs can be identified as the regions (states) whose inter-region contributions to others are significant. In our study, we observe that spatial attention weights do not differ significantly across different samples. These weights represent an overall, consistent spatial relationship among regions and across days. Therefore, in the following analyses, we present both the results aggregated from all test samples as well as the representative test samples.

5.3.1 Local vs. global contributions

To examine the differences between the local (intra-region) and global (inter-region) contributions for forecasting events, we create a contribution graph for each model. As shown in Fig. 8, the orange nodes represent states where the offline events are correctly predicted by the model. The gray nodes represent the states where either the events are not correctly predicted or no event occurred, yet still contribute to forecasting events in other states. For visual clarity, we only show gray nodes having an inter-region contribution greater than a certain threshold (0.01, 0.05 and 0.01 for Charlottesville, Ferguson I and Ferguson II, respectively) to any of the orange nodes. An edge arrow indicates the contribution of forecasting a target state from a source state and the edge weight (encoded by the thickness) reflects the contribution magnitude. Also for visual clarity, we only show edges whose weights are more than a certain threshold, which is 0.05, 0.1 and 0.05 for Charlottesville, Ferguson I and Ferguson II, respectively. For a target state, the self-loop represents the intra-region contribution while other incoming edges represent the inter-region contributions to that state. Note that there might be states where events occurred on multiple days. For such states, we show the average contributions in the graph.

The hierarchical attention mechanism in our ActAttn model enables a systematic way to interpret the intra- and inter-region contributions. The contribution from a source state to a target state (inter-region) on a specific event day is calculated by $(\alpha_{sp} * \alpha_{source})$, where α_{sp} is the attention weight corresponding to the spatial component and α_{source} is the attention weight for the source state in the spatial component, $\mathcal{M}^{sp}$. Similarly, the intra-region (local) contribution can be estimated by $(\alpha_{tem} + \alpha_{sp} * \alpha_{target})$, where α_{tem} is the attention weight corresponding to the (Intra-) temporal component and α_{target} is the attention weight for the target state in the spatial component. As shown in Fig. 8(a), VA has a salient contribution (as a part of global contribution) to forecast the states where the events are correctly

predicted for the Charlottesville case. In other words, social media activity in VA would be a powerful signal for forecasting offline events in the other states. Moreover, CA (mostly), IL and MO can be regarded as hubs, as they contribute more than others to the target states for forecasting events in Ferguson I (Fig. 8(b)). On the other hand, the inter-region contributions from CA and NY to target states are much greater than the other states in Ferguson II (Fig. 8(c)). Note that local (intra-region) contributions (reflected by the self-loop weights) for any target state are higher than the contributions from any other state in all three models. This suggests that local activity still plays a more important role than the activity of any other states. Interestingly, in the case of Charlottesville, the global contribution (the total inter-region contributions of all other states) of a target state is more than the local one, suggesting that the Charlottesville protests have a very distinct spatiotemporal process compared with other the two cases.

5.3.2 The effect of hubs

To further illustrate the hub effect, we select the representative test samples obtained from Texas (TX), Washington (WA) and Illinois (IL), which are correctly predicted events by the Charlottesville, Ferguson I and Ferguson II models, respectively.

In the Charlottesville model, the spatiotemporal attention weights for local and global contributions are 0.458 and 0.542, respectively, meaning that the global part contributes more to forecasting the protest in TX for the given sample. To further analyze the global contribution and hub effect, we visualize the inter-region input (gate) weights and the spatial attention weights as shown in Fig. 9. We observe that Group Lasso regularization selects informative features from only a few states—namely VA, New York (NY), CA and TX (Fig. 9(1a))—and the spatial attention layer further selects VA, CA and NY as hubs (Fig. 9(1b)). VA is the most contributing hub in predicting the protest event for the given test sample from TX. Since the trigger event of the Charlottesville Rally occurred in VA, higher attention weight for VA is the potential indicator that our proposed model is able to model spatiotemporal relationship among the regions successfully for the Charlottesville dataset.

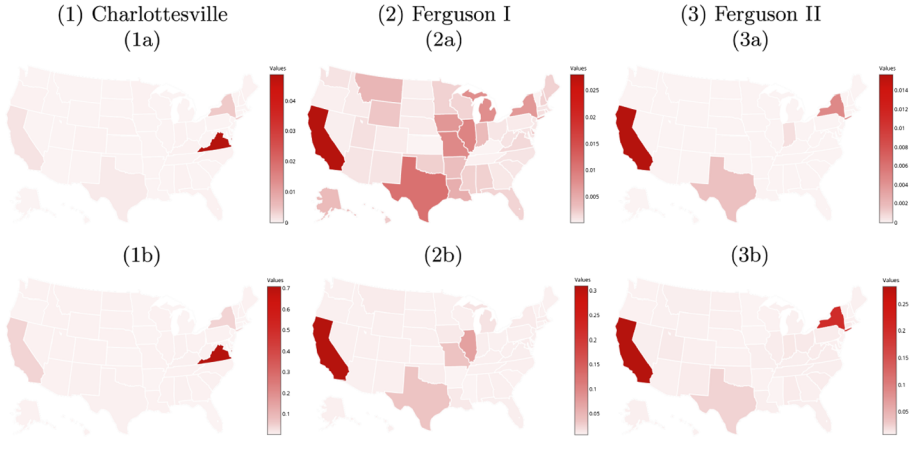


Figure 9 Exploration of global contribution and hub effect. (**a**) Mean absolute values of inter-region input weights across states. (**b**) Attention weights of spatial attention for predicting protests in TX (1b), WA (2b), and IL (3b)

In the Ferguson I model, the spatiotemporal attention weights for local and global contributions are 0.591 and 0.409, respectively. This indicates that locality is more predictive for the given test sample of WA. Spatial attention attends the states CA, IL, Missouri (MO) and TX (Fig. 9(2b)), suggesting the high impact of these states. Ferguson is located in St. Louis, MO where the shooting of Michael Brown happened. It is also very close to the IL border. The reactions to the Ferguson shooting on social media most likely started spreading from these states. CA is an active state where both online (tweet volume) and offline activities occurred much more frequently than other places.

In the Ferguson II model, in predicting the protests in IL, the spatiotemporal attention weights for local and global contributions are 0.576 and 0.424, respectively, for the correctly predicted test sample from IL. As shown in Fig. 9(3a) and Fig. 9(3b), CA and NY are selected by the spatial attention as the most attended regions (among those initially given by the Group Lasso). This suggests that the protest forecasting may be impacted by the heightened social media discussion in these hub states, in relation to, for example, the NYPD shooting of Akai Gurley and the arrest of BLM activists in the Bay Area during the study period.

5.4 Testing predictive power with additional features

While our selection of features is theory-driven, we also consider the possibility of incorporating additional features, which are emerging from the events unfolding, that could help increase the predictive power of the model in a meaningful way. For example, specifically, we consider whether there are keywords utilized by Twitter users to plan, organize, or mobilize protests that may also serve as effective features. Because mobilization activities and activism on Twitter, in most cases, are organized and advocated by Twitter users through hashtags, we focus on identifying the most widely-used hashtags. We analyze the top-k (k = 100) hashtags based on TF-IDF values. We treat each day as a document. We then include these top-100 as additional features to see if they affect forecasting, and analyze the most predictive features.

We assign the ratio of number of tweets that include the hashtag to the total number of tweets at the specific time (day) as the feature value for the corresponding hashtag. According to the results given in Table 4, employing the additional features decreases the performance in terms of both F-score and AUC for all three datasets. Furthermore, we explore the importance of these hashtag features by analyzing the input weights. In all three cases, less than 10% of the features have non-zero weights after Group Lasso regularization, meaning that most of the features do not have any contribution to forecasting events as both intra- and inter-region features. The informative hashtags include: "#theresistance" for Charlottesville; "#ferguson," "#mikebrown" and "# justuceformikebrown" for Ferguson I; and "#ferguson," "# ericgarner," "#tamirrice" and "#fergusondecision" for Ferguson II. However, the weights of these features are much less than the weights of those theory-driven features we first employ in the original model.

Table 4 Forecasting results with and without hashtag features. C.F. stands for content features

	Charlottesville		Ferguson I		Ferguson II	
	F-score	AUC	F-score	AUC	F-score	AUC
Without C.F.	**0.400**	**0.843**	**0.462**	**0.822**	**0.471**	**0.853**
With C.F.	0.308	0.814	0.453	0.815	0.435	0.825

6 Discussion and future work

In this work, we presented an interpretable, predictive model to forecast offline protest events from online activities. We developed a novel deep learning architecture which effectively learns a hierarchical structure of effective features, and at the same time, enables a theory-relevant interpretation. Through extensive experiments, we demonstrated the strength of the proposed model; compared with the baseline methods, our model achieved superior forecasting performance for all movement datasets. It was also more robust with regard to missing data, and consistently outperformed other methods in various early forecasting settings.

Our model not only outperforms existing prediction techniques, but also enables a theory-driven feature selection, together with the differentiation of the intra- and inter-region inputs, allowing us to examine whether these theorized factors are useful in predicting protests as well as how the theoretical framework could help to interpret the model's efficacy and distinct performance across the chosen three threads of protests in a meaningful way. Such an approach could offer insights for further investigations regarding the nature and happenings of protests. Here, we first summarize and explicate whether and how the theory-driven features contribute to forecasting protests. We then discuss the limitations of our work and potential future directions.

6.1 Interpretation of the theory-driven features

First, overall, the greater volumes of tweeting and networking behaviors (including original tweets, replies, and associated content with hyperlinks) had strong predictive power. This result is consistent with prior empirical studies (e.g., [11])—more online discussions may reflect higher public awareness and concern regarding the focal issues and events associated with protests and they opened a cyperspace of social embeddness. Yet, our model allows more differentiating observation and interpretation across protests, in terms of how the social embeddedment was shaped—by messages and interactions within the local state or beyond. For example, we found that *number of reply* played a more significant role only in Charllottesville, suggesting that there may be different natures of how the social embeddedness was created between Charllottesville and Ferguson. Also, *number of URL link* was much more useful in Ferguson II when the tweets came from the local state where the protests happened than when they came from other states.

Second, negative emotions have been studied and theorized to be associated with protests [4, 6], and our results are consistent with this—particularly anger. However, other negative emotions, such as disgust, hate, and fear also stood out, and had distinct predictive power for the Charlottesville counter-protests, Ferguson I and Ferguson II, respectively. Such results, together with our manual inspection of the content of sampled tweets in order to understand what these emotions suggested, also offer insights for future studies in social movements to examine the associations between particular emotions and the nature of protests across contexts.

Third, while one of the operationalizaion of theorized factors, grievance, did not turn out as planned by leveraging the Moral-Laden dictionary, we discovered that the language pattern of negation could be a potential signal of grievance. We discovered in the prediction results that negation (from the LIWC dictionary) could be a good predictor feature for all protest cases, and our manual inspection of the sampled tweets revealed that its semantic meaning could serve as an indicator of grievance. This could be a potential means to identify information of grievance in future relevant studies.

Finally, identity, operationalized by using the *social* category from the LIWC dictionary was able to capture the group identities, and the results showed its predictive power, especially for Charlottesville and Furguson I, but not Furguson II; the second-person pronoun is more predictive in Charlottesville, and the third-person in Ferguson I.

In brief, our model goes beyond indicating that online discussion, including emotional tweets, may help predict offline protests. That point has been studied and widely recognized. Rather, our study offers insights as to where (intra- or inter-) and how (the features were not selected randomly or through unsupervised learning, but theory-driven) the features may offer explanatory power.

6.2 Limitations and future work

There are some limitations in our current work. (1) Our results indicated that considering spatial relationships among the locations increases the performance of forecasting protest events. However, the proposed architecture models the spatial structure irrespective of the locations of events. In other words, it does not differentiate the pairwise relationship between a particular event location and other locations. Future research might consider modeling the relationships between pairs of locations. (2) In the context of forecasting protests or other civil unrest events, data is generally sparse in terms of event occurrences. Events either increasingly happen within a short period after a trigger event, or only occur in particular locations. The data sparsity makes it difficult to learn complex spatiotemporal relationships. Our current model was not specifically designed to tackle this data sparsity issue. (3) In the currently-proposed architecture, the spatial component $\mathcal{M}^{sp}$, which models the spatial relationships over locations, is a complex component. It consists of a set of temporal components for every location, where each component has its own LSTM component. As the number of locations increases, the number of parameters to be learned increases linearly. Although Group Lasso regularization has significantly reduced the complexity of this component, further reducing the complexity of the model would be more desirable.

As part of our future work, we plan to address the aforementioned limitations. In particular, we plan to explore generative models as a solution to overcome data sparsity problem for event forecasting, as well as simplifying the model using weight sharing mechanism.

Acknowledgements
The authors would like to acknowledge the support from NSF #1634944, #1637067, #1739413, and the University of Pittsburgh ULS Open Access Author Fee Fund. Any opinions, findings, and conclusions or recommendations expressed in this material do not necessarily reflect the views of the funding sources.

Abbreviations
BLM, Black Lives Matter; LSTM, Long Short-Term Memory; HMM, Hidden Markov Model; LIWC, Linguistic Inquiry and Word Count; RNNs, Recurrent Neural Networks; LR, Logistic Regression; SVM, Support Vector Machine; GL, Group Lasso; MTFL, Multi-Task Feature Learning; RMTFL, Regularized Multi-Task Feature Learning; CMTFL, Constrained Multi-Task Feature Learning; NN, Neural Network; SGD, Stochastic Gradient Descent; AUC, Area Under Curve; NYPD, New York Police Department.

Availability of data and materials
Data and code are available at https://github.com/picsolab/actattn.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
YRL, WTC, and AME conceived and designed the study. AME conducted the experiments. MY and AL contributed to the data collection and processing. AME, YRL and WTC analyzed and interpreted the results and wrote the manuscript. All authors read and approved the final manuscript.

Author details
[1]School of Computing and Information, University of Pittsburgh, Pittsburgh, USA. [2]Graduate School of Informatics, Middle East Technical University, Ankara, Turkey. [3]Department of Psychology in Education, School of Education, University of Pittsburgh, Pittsburgh, USA.

Endnotes
[a] Keywords include: Charlottesville, KKK, Ku Klux Klan, Klansman, Klansmen, Nazi, Nazism, racism, racist, supremacy, supremacist, supremacists, #Charlottesville, #domesticterrorism, # FireBannon, #WhiteSupremacist, #WhiteSupremacists.
[b] https://elephrame.com/.
[c] While the tweets for Charlottesville and Ferguson were collected separately using different collection methods, the information about protest events was collected from the same data source—the Elephrame website. As we mainly focus on the spatiotemporal patterns of the offline protest events, the difference in terms of methods used for collecting tweets will not significantly impact our results and interpretation.
[d] https://www.britannica.com/topic/list-of-cities-and-towns-in-the-United-States-2023068.
[e] The AUC of the best model (>0.82) suggests it is possible to rank-order or filter the states where protest events are likely to happen with reasonable accuracy.

Publisher's Note

Received: 4 July 2018 Accepted: 24 January 2019 Published online: 14 February 2019

References
1. Snow DA, Soule SA, Kriesi H (2008) The Blackwell companion to social movements. Wiley, New York
2. Valenzuela S (2013) Unpacking the use of social media for protest behavior: the roles of information, opinion expression, and activism. Am Behav Sci 57(7):920–942
3. Theocharis Y, Lowe W, van Deth JW, García-Albacete G (2015) Using Twitter to mobilize protest action: online mobilization patterns and action repertoires in the occupy wall street, indignados, and aganaktismenoi movements. Inf Commun Soc 18(2):202–220
4. Van Stekelenburg J, Klandermans B (2013) The social psychology of protest. Curr Sociol 61(5–6):886–905
5. Klandermans B, van Stekelenburg J (2013) The political psychology of protest. Eur Psychol 18(4):224–234
6. Goodwin J, Jasper JM (2006) Emotions and social movements. In: Handbook of the sociology of emotions. Springer, Berlin, pp 611–635
7. González-Bailón S, Borge-Holthoefer J, Rivero A, Moreno Y (2011) The dynamics of protest recruitment through an online network. Sci Rep 1:197
8. Conover MD, Ferrara E, Menczer F, Flammini A (2013) The digital evolution of occupy wall street. PLoS ONE 8(5):64679
9. Conover MD, Davis C, Ferrara E, McKelvey K, Menczer F, Flammini A (2013) The geospatial characteristics of a social movement communication network. PLoS ONE 8(3):55957
10. He J, Hong L, Frias-Martinez V, Torrens P (2015) Uncovering social media reaction pattern to protest events: a spatiotemporal dynamics perspective of ferguson unrest. In: International conference on social informatics. Springer, pp 67–81
11. De Choudhury M, Jhaver S, Sugar B, Weber I (2016) Social media participation in an activist movement for racial equality. In: ICWSM, pp 92–101
12. Qi H, Manrique P, Johnson D, Restrepo E, Johnson NF (2016) Open source data reveals connection between online and on-street protest activity. EPJ Data Sci 5(1):18
13. Ferguson unrest. https://en.wikipedia.org/wiki/Ferguson_unrest. Accessed: 2018-04-01
14. Unite the Right rally. https://en.wikipedia.org/wiki/Unite_the_Right_rally. Accessed: 2018-04-01
15. Zhao L, Sun Q, Ye J, Chen F, Lu C-T, Ramakrishnan N (2015) Multi-task learning for spatio-temporal event forecasting. In: Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 1503–1512
16. Zhao L, Wang J, Chen F, Lu C-T, Ramakrishnan N (2017) Spatial event forecasting in social media with geographically hierarchical regularization. Proc IEEE 105(10):1953–1970
17. Scardapane S, Comminiello D, Hussain A, Uncini A (2017) Group sparse regularization for deep neural networks. Neurocomputing 241:81–89
18. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
19. Chung WT, Lin YR, Li A, Ertugrul AM, Yan M (2018) March with and without feet: the talking about protests and beyond. In: International conference on social informatics. Springer, pp 134–150
20. Panagiotou N, Zygouras N, Katakis I, Gunopulos D, Zacheilas N, Boutsis I, Kalogeraki V, Lynch S, O'Brien B (2016) Intelligent urban data monitoring for smart cities. In: Joint European conference on machine learning and knowledge discovery in databases. Springer, Berlin, pp 177–192
21. Teng X, Yan M, Ertugrul AM, Lin YR (2018) Deep into hypersphere: robust and unsupervised anomaly discovery in dynamic networks. In: International joint conference on artificial intelligence.
22. Gerber MS (2014) Predicting crime using Twitter and kernel density estimation. Decis Support Syst 61:115–125
23. Korkmaz G, Cadena J, Kuhlman CJ, Marathe A, Vullikanti A, Ramakrishnan N (2015) Combining heterogeneous data sources for civil unrest forecasting. In: 2015 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM). IEEE, New York, pp 258–265
24. Korolov R, Lu D, Wang J, Zhou G, Bonial C, Voss C, Kaplan L, Wallace W, Han J, Ji H (2016) On predicting social unrest using social media. In: 2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM). IEEE, New York, pp 89–95

25. Cadena J, Korkmaz G, Kuhlman CJ, Marathe A, Ramakrishnan N, Vullikanti A (2015) Forecasting social unrest using activity cascades. PLoS ONE 10(6):0128879
26. Ning Y, Muthiah S, Rangwala H, Ramakrishnan N (2016) Modeling precursors for event forecasting via nested multi-instance learning. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 1095–1104
27. Ramakrishnan N, Butler P, Muthiah S, Self N, Khandpur R, Saraf P, Wang W, Cadena J, Vullikanti A, Korkmaz G et al (2014) 'Beating the news' with EMBERS: forecasting civil unrest using open source indicators. In: Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 1799–1808
28. Zhao L, Chen F, Lu C-T, Ramakrishnan N (2015) Spatiotemporal event forecasting in social media. In: Proceedings of the 2015 SIAM international conference on data mining. SIAM, Philadelphia, pp 963–971
29. Zhao L, Wang J, Guo X (2018) Distant-supervision of heterogeneous multitask learning for social event forecasting with multilingual indicators. In: AAAI
30. Chung C, Pennebaker JW (2007) The psychological functions of function words. In: Social communication, pp 343–359
31. Ma J, Gao W, Mitra P, Kwon S, Jansen BJ, Wong K-F, Cha M (2016) Detecting rumors from microblogs with recurrent neural networks. In: IJCAI, pp 3818–3824
32. Tuor A, Kaplan S, Hutchinson B, Nichols N, Robinson S (2017) Predicting user roles from computer logs using recurrent neural networks. In: AAAI, pp 4993–4994
33. Hu W, Singh KK, Xiao F, Han J, Chuah C-N, Lee YJ (2018) Who will share my image? Predicting the content diffusion path in online social networks. In: Proceedings of the eleventh ACM international conference on web search and data mining. ACM, New York, pp 252–260
34. Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align and translate. Preprint. arXiv:1409.0473
35. Denil M, Bazzani L, Larochelle H, de Freitas N (2012) Learning where to attend with deep architectures for image tracking. Neural Comput 24(8):2151–2184
36. Zhao L, Hu Q, Wang W (2015) Heterogeneous feature selection with multi-modal deep neural networks and sparse group lasso. IEEE Trans Multimed 17(11):1936–1948
37. Zhu W, Lan C, Xing J, Zeng W, Li Y, Shen L, Xie X et al (2016) Co-occurrence feature learning for skeleton based action recognition using regularized deep LSTM networks. In: AAAI, vol 2, p 8
38. de Albornoz JC, Plaza L, Gervás P (2012) Sentisense: an easily scalable concept-based affective lexicon for sentiment analysis. In: LREC, pp 3562–3567
39. Graham J, Haidt J, Nosek BA (2009) Liberals and conservatives rely on different sets of moral foundations. J Pers Soc Psychol 96(5):1029
40. Freelon D, McIlwain CD, Clark MD (2016) Beyond the hashtags: #ferguson, #blacklivesmatter, and the online struggle for offline justice
41. Kingma D, Ba J (2014) Adam: a method for stochastic optimization. Preprint. arXiv:1412.6980