Identifying violent protest activity with scalable machine learning*

Lefteris Anastasopoulos †
Assistant Professor
School of Public and International Affairs
Georgia Informatics Institute
University of Georgia

Jake Williams‡
Assistant Professor
Department of Information Science
College of Computing and Informatics
Drexel University

August 28, 2016

Abstract

The outbreak and frequency of violent protest activity since 2010 has been a cause for alarm among policy makers and the public at large and has renewed interest in the study of violent forms of protest action. Until recently, the study of violent protest action, and indeed protest action in general has been limited to case studies (Tilly 1988; Tilly and Tarrow 2015), simulation studies (Epstein 2002) and newspaper accounts (Earl et al. 2004). With the widespread use of social media websites such as Twitter and Facebook as means of protest mobilization along with innovations in high dimensional statistics and machine learning researchers are now able to collect large and geographically diverse data for studying protest activity (Barberá et al. 2015; Metzger et al. 2016; Gerbaudo 2012; Tucker et al. 2014). In this paper, we build a series of scalable machine learning algorithms and software which jointly leverage spatial and textual data to identify violent and peaceful protest activity using English language Tweets. We then use our classifier to demonstrate how our software can be used by researchers to construct databases which measure violent and peaceful forms protest activity at fine-grained levels of time and geography and explore relationships between Census demographics and protest activity during the Ferguson protests in November 2015. Finally, we explore how linguistic and spatial features distinguish peaceful from violent forms of collective action.

*All errors are our own.

†ljanastas@uga.edu, http://scholar.harvard.edu/janastas

‡jakerylandwilliams@gmail.com, http://people.ischool.berkeley.edu/%7Ejakeryland/

1 Introduction

In the United States and in Europe, the rise to power and popularity of "law and order" conservative political figures such as Donald Trump in the US are in no small part due to reactions by citizens of these countries to perceived threats stemming from increases in immigration and an increase in the frequency of violent protests in these nations and others around the world (Anastasopoulos 2015; May 2016; Sørensen 2016). In 2016 in the US alone, city names such as Milwaukee, Dallas and Chicago have become synonymous with violent protest activity as large-scale urban riots, similar to those witnessed in the late 1960s and 1970s in the United States, appear to be resurgent (Carter 1986; Clarke and Egan 1972; Collins and Margo 2007).

As violent protests continue to erupt in cities across the world, there is a growing need to understand the conditions that lead to them. Violent protest activity not only poses significant problems for individuals and communities directly affected by them, but also diminishes the legitimacy of causes associated with them (Huet-Vaughn 2013; Porta and Tarrow 1986), thus hindering needed social change. For researchers, a tool which can identify violent and peaceful protests can provide a rich source of data which will expand the scope of knowledge and understanding of modern collective action and overcome the inherent data availability limitations which have restricted the study of protests to either detailed case studies (Tilly 1988; Tilly and Tarrow 2015) or news media sources (Earl et al. 2004). From a public safety perspective, citizens would benefit from a tool which could warn them about areas in which violent activity is currently occurring or where it is likely to occur so that these areas can be avoided. Additionally, by identifying where violent activity is likely to take place, preventative measures can be put into action by law enforcement and others which could potentially save lives.

In this paper, we design a series of original scalable machine learning algorithms which are able to identify violent and non-violent protest activity using a massive geocoded Twitter database. Our available data includes all geo-coded Tweets between April, 1st,

2014 and April 30th, 2015. Using our software, we first explore the linguistic and spatial features of Tweets that distinguish Tweets unrelated to protest activity from those related to violent and non-violent forms of collective action. We then demonstrate how our software can be used to build databases which contain metrics of protest activity. Finally, using some sample data from this database, we explore relationships between Census tract level demographics and protest activity in the United States during the month of the Ferguson protests in November of 2014.

2 Protest Activity and Mobilization

Tilly (1978) notes that the analysis of collective action "has five big components: interests, organization, mobilization, opportunity and collective action itself." Interests are benefits received by participants in collective action. Organization is the structure of groups which allow participants to achieve their goals. Mobilization includes resources groups require to organize their activities. Opportunity is related to the groups interests and those of the world around them and collective action encompasses the forms of action taken to achieve a collective goal, of which protest is a subset. For our purposes, we are concerned with leveraging the fact that social media is currently one of the major protest mobilization tools used around the world but before delving into the details of our model, it is useful to understand how protest activity has been defined by scholars in the past and some of the mobilization techniques which have evolved over time.

At a very basic level a "protest" is well understood by readers of modern English as an activity engaged in by a group of individuals to effect political change of some sort. Early theories of protest agreed that those who engage in protest activity often lack the political resources to effect change through other means (Lipsky 1968). As James Q. Wilson famously noted, protest is a "problem of the powerless" (Wilson 1961). This dictum can be observed from a brief exploration of protest activity in the United States in the latter part of the 20th century. During the civil rights movement, protest was one

of the most powerful tools available to disenfranchised African-Americans seeking to end policies which explicitly and discriminated against them.

While the varieties of protest causes throughout the modern era have ranged from the overthrow of King Louis XVI during the French Revolution to transgender bathrooms in 2016, the one constant required for a successful protest to occur is a mobilization effort which can effectively gather large groups of people in one or a few central locations. Prior to the invention of the internet, protest mobilization occurred primarily through print news sources, flyers and telephones. While the invention of the telegraph, radio and television in the 20th century all had a great deal of potential as mobilization tools, these fora were available mostly to government entities and businesses with a significant amount of money and resources (Tarrow and Tollefson 1994). Indeed, it was not until the Internet age and availability of cheap internet service that marginalized individuals, those most likely to participate and benefit from protests, were able to easily discover and participate in local protest activity and mobilization. Since social media outlets such as Twitter were established, however, these sites have become organizational hubs of protest activity and have likely changed the nature of protest activity as well (Gerbaudo 2012).

Although links between protest mobilization tools and protest violence are presently unclear, the literature on the evolution of protest violence points to cycles of transition from peaceful forms of civil disobedience to more violent forms of protest . To explain the vicissitudes of violent and peaceful protest activity, scholars from a wide range of fields have explored the micro- and macro-level motivations of decisions to engage in violent and non-violent protest activity. Below we provide a brief description of the literature in these areas (Moore 1998; Porta and Tarrow 1986).

2.1 Micro Level Models of Violent Protest

Peaceful and violent forms of protest activity are not opposing tactics. Rather they can be seen simply as difference forms of "negative inducements" that protestors must provide

in order to acheive their goals (Lipsky 1968). Game theoretic models and simulations focused on the behaviors of individual agents and their relationships to the collective argue that whether a protest is peaceful or violent depends upon a well defined number of factors. Below we describe an updated version of one of the most well known of these models: the Brookings Model of Civil Violence (Epstein 2002). In this model, motivations behind protest activity are understood as the product of three major factors: levels of perceived grievance **G**, probability of arrest **P** and risk aversion.

Grievance **G**, which is itself a product of hardship **H**) and legitimacy **L** represents the seriousness of the grievance of the protester. Hardship represents the difficulties that the protester faces in their daily lives as the result of poverty, privation, etc. and legitimacy (**L**) is a measure of the perceived legitimacy of a regime, central authority or group that the protesters are targeting[1]. Putting all of these terms together protest grievance **G** of an agent i is:

$$G_i = H_i(1 - L_i) \tag{1}$$

In Equation 1 grievance is thus a linear function of personal hardship and the perceived "illegitimacy" $(1-L)$ of the group toward which the protest is being directed. In the case of anti-police protests such as those that broke out in 2014 in Ferguson, MO after the death Michael Brown and in Baltimore, MD after the death of Freddie Gray, it is clear that grievance levels against the police were high as those that participated in the violent protests had both high degrees of hardship, as measured by poverty and high degrees of perceived beliefs that the police departments were not legitimate if we are to go by media accounts.

[1]The Brookings Model was originally designed to address civil conflict at a national level but can be easily adapted to represent civil violence at the a local level. In the case of protest for reasons of civil disobedience, for example anti-police protests, **L** can be thought of as the perceived legitimacy of the police force.

Interestingly, how to measure **H** and **L** was not discussed in any detail in (Epstein 2002) or subsequent works employing this model, but it is worth discussing in the context of using social media to study protest activity. While one can produce relatively objective measures of hardship based on annual income, debt, etc. the legitimacy of a police force, for example, depends more on experience with members of the force and how the police are portrayed in media accounts and by influential public figures. Thus, we would add to this model that legitimacy is, in turn, a function of actual experience ϵ, with the external authority and perceptions of the external authority as portrayed by a third-party such as a major news source or through social media, μ. Thus, legitimacy and grievance in more modern protests is defined by us as:

$$L_i = f(\epsilon_i, \mu) \tag{2}$$

$$G_i = H_i(1 - f(\epsilon_i, \mu)) \tag{3}$$

From the equations above, the role media plays in determining grievance levels is clearer. Because media, either via news or social media, can manipulate the illegitimacy of the external authority that groups are protesting against, they can significantly increase the levels of grievance among protesters.

The probability of arrest by police **P** in this model and the risk aversion factor of the protester, R are also key components. The probability of protest, is a function primarily of the cop C to protester A ratio C/A scaled by a constant k:

$$P = 1 - exp[-k(C/A)] \tag{4}$$

In equation 4 the probability of arrest is directly related to the cop to protester ratio such that it increases as the number of police increase holding protesters constant and decreases as the number of protesters increase holding the police force constant. Finally,

risk aversion **R** is the amount of risk aversion that a protester has.

Combining all of these elements into a decision rule, violent protest breaks out when:

$$G - RP > 0, \tag{5}$$

or when the grievance is greater than the agent's risk aversion and the probably of arrest.

While the Brookings model is a somewhat oversimplified description of the factors that lead to the outbreak of violent protest activity, it provides a useful framework for thinking about the most basic elements of violent protest which can easily apply to real life situations. For example, the outbreak of violent protests in Oakland, CA soon after the acquittal of Darren Wilson, a white officer that killed Michael Brown, an African-American teenager, can be understood in the context of this model. Oakland has pockets of extreme poverty and a high crime rate. In the context of the Brookings model this translates to high levels of hardship and and low levels of legitimacy due to frequent encounters with the police. After the verdict, grievance levels measurably skyrocketed as a wave of outrage spread across social media and violent protests soon broke out in Oakland.

The Brookings model was subsequently picked up by scholars who used it in large-scale agent based modeling simulations to study topics related to ethnic segregation and violence in the Middle East (Weidmann and Salehyan 2013), urban violence (Bhavnani et al. 2014) and more complicated, dynamic simulations on related to violent protest (Torrens and McDaniel 2013).

2.2 Macro Level Models of Violent Protest

Macro level models of violent protest focus on the ebbs and flows of protest activity over longer periods of time and attempt to build structural explanations of cycles of violent and non-violent protest activity (Meirowitz and Tucker 2013). Tarrow (1989),

for example, notes that the "largest current problem in collective action research" is to explain "...not why people periodically petition, strike, demonstrate, riot, loot and burn, but rather why so many of them do so at particular times in their history, and if there is a logical sequence to their actions." To accomplish this goal, scholars in this field have focused on the context that protests have taken place in and linked historical contingency to violent protest activity using case studies (Seferiades and Johnston 2012; Tarrow and Tollefson 1994; Tilly 1988; Tilly and Tarrow 2015) or panel data sets compiled from newspaper coverage (Huet-Vaughn 2013)

One of the major models in this area was proposed by (Tarrow and Tollefson 1994) who argued that contentious politics, which includes protest in both violent and non-violent forms "emerges in response to changes in political opportunities and constraints." Violence, according to (Tarrow and Tollefson 1994), is a form of "public performance" in which actors challenge status quo policies or actions committed by the state. As mentioned above, while it is one of the most visible forms of collective action, it is one of the least effective.

3 Framework for Measuring Social Action

We develop a scheme for measuring social action from textual data that hinges on two defining characteristics. In particular, we are interested whether any represented actions are peaceful or forceful and collective or singular. Under this framing, if an action is identified it can fall in one of four categories defined by any pairing from the two characteristics. For example, an expression of empathy or support from an individual would fall under singular peace, while a report of looting or a police blockade would fall under collective force. This places violent social actions under the collective and singular force categories, while leaving room for the identification of other actions that would not necessarily be considered violent, e.g., the protest marches down streets and freeways that ensued following Darren Wilson's trial. However, as collective protest actions can

also be peaceful (e.g., vigils, chanting, and revered silence), we'll see protest activity split across two categories under our model.

4 Data

To explore the measurement of social action under our model, we utilize a database of over 600 million high-precision geographically-tagged tweets from the Twitter social network collected over April 2014–April 2015, which most notably covers the Black Lives Matter (BLM) movement extensively. This dataset was collected from Twitter's public (spritzer) API, and is of particular importance for Twitter's policies around location tagging at the time of collection. At the time, when a user opted in for location tagging from a mobile device, a tweet sent would be accompanied by high-precision latitude and longitude coordinates. Since then, Twitter went into an agreement with Foursquare that resulted in the adjustment of their system to front the option of soft-locations, which users specify (for example, one could set their location to Philadelphia and then go on to tweet from anywhere else in the world, with tweet meta-data always listed as Philadelphia). Since the rate at which tweets were geo-tagged was approximately 1% at the time of collection, and the (1% stream) public API was restricted to *only* location-tagged tweets, it is arguable that this collection represents the majority of geo-tagged tweets from the period of collection.

In addition to the geo-tagged Twitter database, we also utilize the meta-data held in Associated Press images labeled with the term "protest" from the same time period. This data informs us of times and locations of protest events from around the world, spanning the same time period as the Twitter database, for the coding of social action (assuming this sample to be particular potent with representations of the actions of interest) (*Associated Press (AP) Image Database* 2016).

5 Methods

Using a small portion (18,000, or approximately 10%) of the tweets sampled from the A.P.-image identified protest times and locations, in addition to all tweets from Alameda county California on the night of Nov. 24th, 2014 (since one of the authors was impinged at this time and location by a protest), we go about coding tweets individually for the presence of the four types of social action (collective/singular peace/force). In particular, we accept that tweets may represent any number of the four types of action. From the total 22,625, a breakdown of the positively coded tweets is give in Tab. 1.

We then use the coded data as input for binary naïve Bayes classifiers, which, for each of the four action types, run in parallel. We modify standard naïve Bayes classifiers with an enhanced input feature space. In particular, instead of using relying on space-delimited "words" as predictive features for classification, we utilized a recently developed (Williams 2016) multiword expression segmentation method to construct distributions of words and phrases of one or more words. Using distributions of phrases has been shown (Williams et al. 2016) to better support the naïve, independence assumption in the classification method.

In addition to improving the Bayes classifier used in our experiments, the usage of phrases as features allows for greater interpretability of classifications. The naïve Bayes classifier has an advantage of being explorable, as a white box method that can be opened, to show the features most relevant to classifications. In particular, looking at a document as a bag of phrases $d = \{w_1, w_2, \cdots, w_N\}$, counted with frequencies $\{f(w_1), f(w_2), \cdots, f(w_N)\}$, their impact on the Bayes classification is largely due to the likelihood function, $\mathcal{L}$ (determined in training), often computed as a sum of logarithms:

$$\log P(c \mid d) \approx -\log P(c) - \sum_{i=1}^{N} f(w_i) \log \mathcal{L}(w_i|c),$$

where c is the class presence (positive/negative), and the terms are negated for an en-

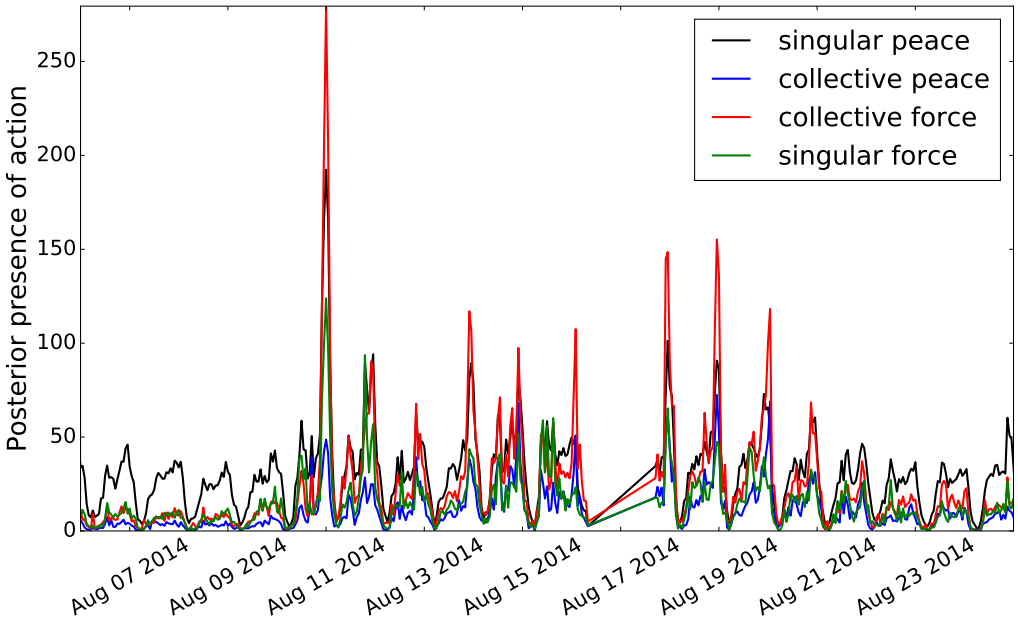

Figure 1: Time series showing the total presence of social action types in Ferguson, MO over several weeks in August, 2014. The presence of each action type is determined by a naïve Bayes classifier, and measured as the sum of posterior probabilities over all tweets from each hour in the plotted span of time.

tropic framing. If a practitioner wishes to understand why, for example, a document was classified as positive (c_{pos}) over negative (c_{neg}), the difference can be seen to be:

$$-(\log P(c_{\text{pos}}) - \log P(c_{\text{neg}})) - \sum_{i=1}^{N} f(w_i)(\log \mathcal{L}(w_i|c_{\text{pos}}) - \log \mathcal{L}(w_i|c_{\text{neg}})),$$

which affords a ranking of the features by the (absolute) terms of the sum:

$$f(w_i)\,|\log \mathcal{L}(w_i|c_{\text{pos}}) - \log \mathcal{L}(w_i|c_{\text{neg}})|$$

that we display as a vertical bar plot, called a phrase shift (see Figs. 2 and 3).

Ferguson, August 11th, 2014 (12 AM, GMT)

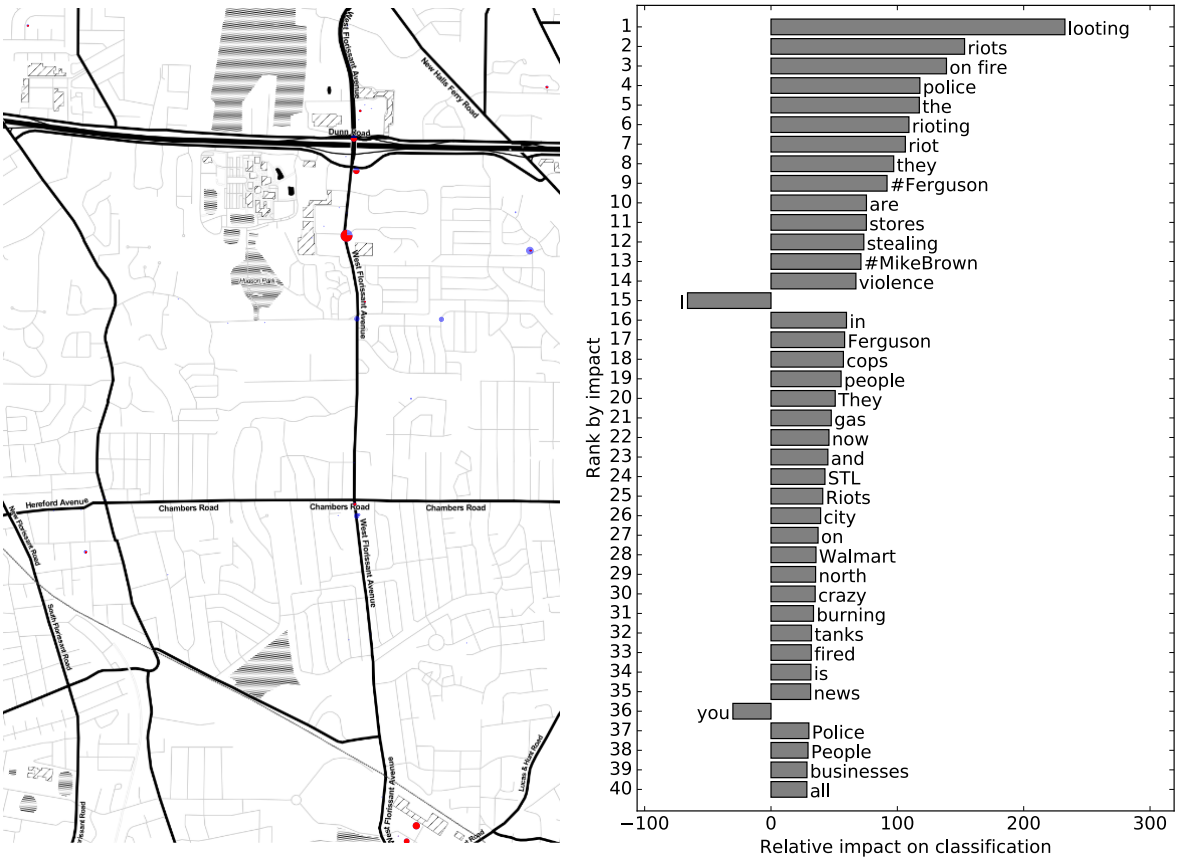

Figure 2: Left. Map of Ferguson, MO depicting clusters of collective force activity over one hour around 12 AM, on August 11th. The size of each cluster-circle represents the area from which tweets emerged (not the number of tweets contained), and the portion of each circle colored red indicates the portion of tweets classified to represent the collective force action. Right. A phrase shift showing the most impactive features present in all tweets classified as being representative of collective force. Phrases on the right pull the classifier toward a positive classification, and phrases on the left pull the classifier towards a negative classification.

6 Experiment and results

We examine the performance of our classifier by performing a tenfold cross-validation on the coded tweets data set. The results of this validation are recorded in Tab. **??**. Treating the Bayes posterior probability as a tunable threshold for classification, we measure precision and recall, and optimize the threshold probability over F_1 to tune each given

Ferguson, August 18th, 2014 (11 PM, GMT)

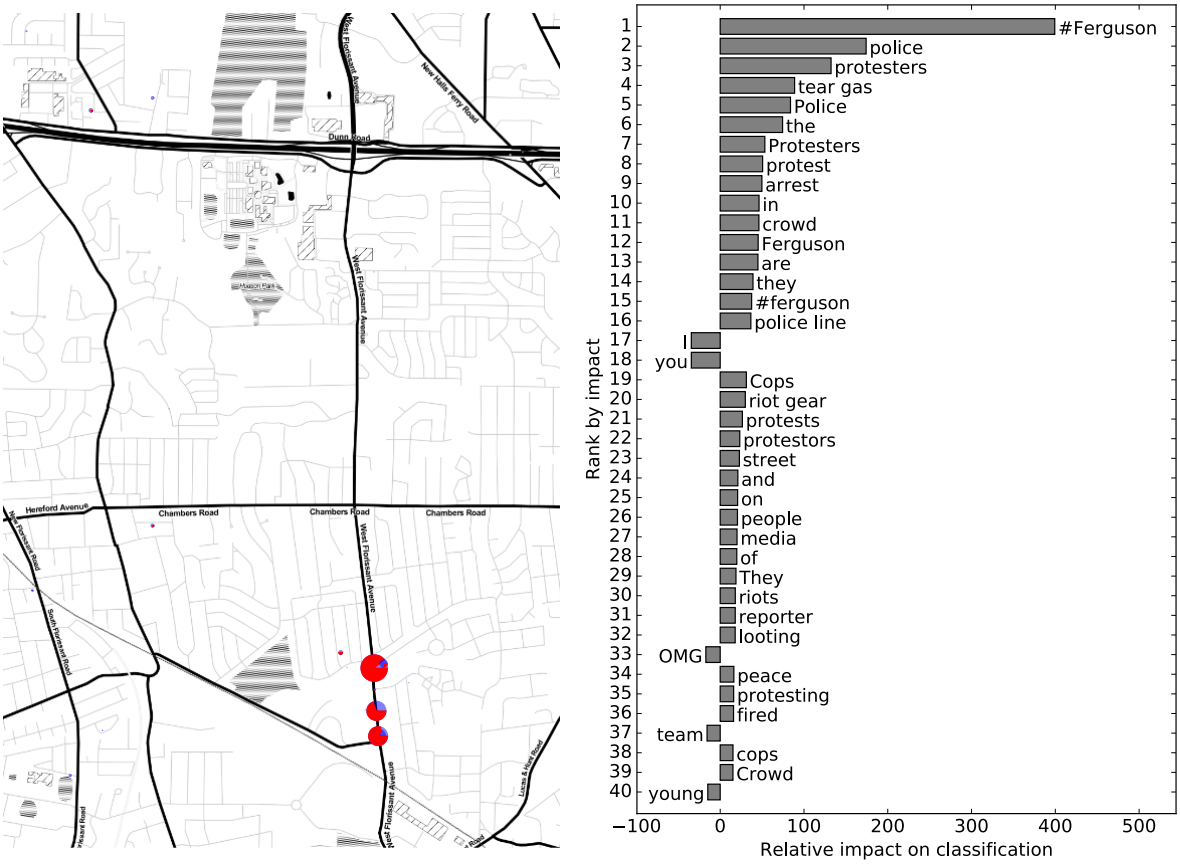

Figure 3: Left. Map of Ferguson, MO depicting clusters of collective force activity over one hour around 11 PM, on August 18th. Right. A phrase shift showing the most impactive features present in all tweets classified as being representative of collective force. Note: points and bars represent analagous quantities to those in Fig. 2.

classifier. Observing these results, we see that collective force is, individually, the best predicted action type. This is encouraging, as collective force often represents the most serious actions. While classifier performance at predicting collective peace and singular is lower, we do see that the most prevalent type of action, singular peace, is predicted well. When the classifiers are collapsed to less-specific types of action (Collective, Singular, Peace, and Force) performance decreases from the best cases (singular peace and collective force), but when all action types are combined (All), we see a significant performance improvement in all measures.

Hong Kong, September 29th, 2014 (11 AM, GMT)

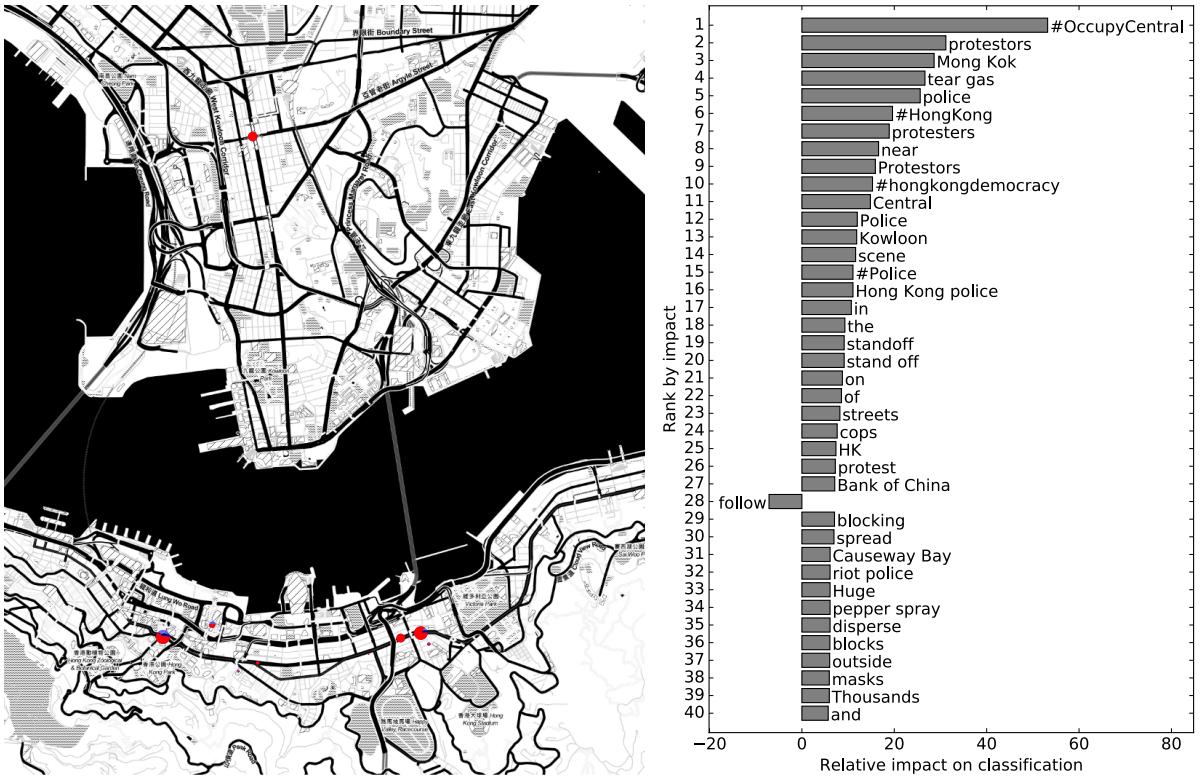

Figure 4: Left. Map of Hong Kong depicting clusters of collective force activity over one hour around 11 AM, on September 29th. Right. A phrase shift showing the most impactive features present in all tweets classified as being representative of collective force. Note: points and bars represent analagous quantities to those in Fig. 2.

To exhibit the manner in which our classifier might be used, we apply our trained classifier to data from outside of training, taken from Ferguson, MO. during the initial wave of protest activity, over August of 2014. In Fig. 1 we plot a time series of this period, showing the abundance of the four types of social action, as measured by the sum of posterior probabilities of all tweets under the application of our classifiers. Here, it can be seen that the largest spikes occurred on the first night of protesting (Aug. 10th), and continued through for approximately two weeks (noting a gap in our data on the 17th). While singular peace (black line) exhibits a substantial, periodic signature even under normal circumstances (the discourse it covers is regular and common), collective force (red line) emerges aberrantly during the protest events, overshadowing the presence of

the other action types.

Taking a closer look at the presence of collective force for some of the larger spikes, we zoom in to a map of the first night of protesting in Fig. 2 (left), and plot clusters of tweets with the positive classifications represented as proportional areas. Here, we can see a larger cluster just south of the freeway, on West Florissant Avenue, which corresponds to the time and location of the burning of the QuikTrip convenience store and gas station (set to fire by protestors). This action is actually hinted at in the phrase shift (bar plot, right), by terms such as "burning" and "on fire." While the first night of protesting was violent and unexpected, the actions that took place were spread out, and involved fewer mass confrontations with the police, which later became more militarized. This change is quite noticeable in Fig. 2, which shows concentrated action on the night of August 18^{th}, when protesters and police clashed violently further south along West Florissant Avenue. We are once again able extract detailed information as to the nature of events at this point through the phrase shifts, which show impactive terms like "police line," "riot gear," and "tear gas."

We additionally present the result of our model's application to the Hong Kong democracy protests that lasted for approximately two months in the fall of 2014. On the map in Fig. 4, we see clusters of collective force activity at the three main protest sites, Admiralty, Causeway Bay, and Mong Kok. Tactics similar to those reported by Twitter users during the Ferguson protests were employed by the Honk Kong police as well, as is indicated by a phrase shift (Fig. 4, right) that shows highly-impactive phrases such as "tear gas," "stand off," and "riot police." So for the collective force category, we see a large degree of accord in the lexical features that indicate the presence of the action (which we quantify below, in Tab. 2), indicating the possibility of applicability to out-of-domain data, and future events.

7 Building a Protest Activity Database

While the literature discusses many aspects of protests one of the central political questions surrounding protest activity involves its effects on public opinion (Dunaway, Branton, and Abrajano 2010; Wallace, Zepeda-Millán, and Jones-Correa 2014; Branton et al. 2015) and its ability to influence policymakers. The effects of protest activity on public opinion are especially important to understand because the former typically influences the latter. To construct measures of protest activity researchers have almost exclusively relied on compiling databases of newspaper coverage of protest activity (Earl et al. 2004). While these databases arguably capture some of the most impactful protest activity, they have limited the study of protests to those picked up by media sources which have resulted in an understanding of the effects of protest activity mediated entirely by the news media's willingness to cover them and are limited in terms of time and geography.

Here, we demonstrate how our software can be used to build a database of measurable protest activity at the county level in the United States. Using our software, we explore patterns of protest activity across the United States shortly after the grand jury acquittal of officer Darren Wilson in Ferguson, MO on November 24th and construct measures of protest activity using 3.5 million classified Tweets.

Figures 5, 6, 7, and 8 are maps containing measures of violent, peaceful and overall protest activity across the United States and within Missouri after the acquittal of Darren Wilson on November 24th, 2014. This data demonstrates the potential for our software to build databases to measure protest activity at relatively low spatial and temporal levels.

8 Discussion

While the social events we have exhibited for Ferguson and Hong Kong are strong examples of the types of action we have sought to model, there are in fact many other examples contained within the 600 million Tweet data set. Many social events are known

to have occurred over this time period, but there are likely many more, smaller events, unknown though traditional media outlets that could be detected in the data with our model of action. An even larger Twitter stream (10% bandwidth) is already being processed in real time for positive and negative affect (i.e., happiness), whose collective instrumentation is referred to as the Hedonometer (Dodds et al. 2011, 2015), and is annotated by hand for it's ability to indicate strongly positive and negative events, temporally. There have also been numerous recent works focused on the automatic detection of events from social media streams in general, such as (Sayyadi, Hurst, and Maykov 2009; Aggarwal and Subbian 2012; Nurwidyantoro and Winarko 2013; Zhou and Chen 2014; Dong et al. 2015), to name a few. The comprehensive nature and extensive coverage (of an event-filled year) offered by the 600 million tweet data set (which was developed by the creators of the Hedonometer) makes it an exciting candidate for further coding, as a training and validation dataset for this community of research. While our model simply identifies the presence of different types of social action, it will be important to comprehensively collate these actions into different social events, for our continued research and use of other researchers.

The information we have depicted in Figs. 1, 2, and 3 serve as an example of the diagnostic utility that our model could provide. A logical subsequent development would seek to build this out as an explorable online tool, so as to enable lay observers to see an understand the patterns of social action in real time. However, there are substantial obstacles to developing a real-time tool. As mentioned, the data taken from Twitter (on which an real-time explorable tool would rely) constitute a sample of data possessing geographic precision like no other, that the current data being produced do not posses. However, high-precision geographic location reporting is still an option that may be elected by users (though more buried in the App.), and if a sufficiently persuasive outreach program were undertaken, it is conceivable that a critical mass of users could be engaged. Also, while the (1% bandwidth) public API has shown to be sufficient (if rich with geo-coding) in quantity for illustrative distributed reporting, a larger stream of geo-coded

data would not doubt improve the quality of or model, and any inferences from it that we wish to draw.

We do see other clear ways for improving our model. Of the data sampled with the support of the AP images, we have only completed the coding of approximately 10%. Certainly, completing the coding of the remaining portion would result in the better training of our model. Incorporating these other data into training would undoubtedly improve the performance of our model, in particular with the range of other language associated with social actions that is more rare, and not covered by the smaller sample used. This would likely improve our model's performance when applied to real-time, out-of-domain data that represent emergent events, which are ultimately of the utmost interest and importance.

While we can't quantify our model's performance in application to real-time data, we can hint at its performance on out-of-domain data, by separating known distinct events in the training data. In addition to the BLM movement, the coded data significantly cover a portion of the Hong Kong democracy protests. Using the data from Hong Kong for testing, and all other data for training, we see somewhat different results (see Tab. 2), and generalize to note some potential challenges with using this data (that only covers a limited set of events) to build a classifier and apply it to data representing unknown and unforeseen events. First, the subject matter (democracy) from the Hong Kong tweets is very different from that of the BLM movement (institutionalized racism), making the discourse present in the singular peace testing tweets largely unrelated to that from training. As a result, this previously predictable category now exhibits substantially decreased performance. Furthermore, of the nearly 900 tweets coded from Hong Kong, only 6 were found to be representative of singular force, so there is essentially nothing to predict for this category. However, for the collective actions we see better numbers, especially for collective force. This is likely as a result of the similar collective tactics employed on both sides of both movements (e.g., blockades, non-lethal pacification, etc.). When the different action type are collapsed, we see more and more performance improvements, indicating

that the collapsed categories may be the most reliable. However, since the Hong Kong tweets are actually part of the training of the overall classifier, we note that the performance of our model when applied to real-time data will likely be significantly better than that reported in Tab. 2, and importantly, for the most serious type of action—collective force—our results in performance are largely upheld. Furthermore, as we continue with the coding of data our coverage of the language surrounding social action will increase, likely improving our model's performance here, too.

Action	No. pos. codings	Posterior prob.	Precision	Recall	F_1
Collective force	795	0.08	74.08	76.01	74.94
Collective peace	474	0.78	51.92	55.25	53.29
Singular force	351	0.92	57.39	41.19	47.38
Singular peace	1,823	0.85	73.52	67.61	70.38
Collective	1,116	0.79	74.54	68.80	71.48
Singular	1,951	0.87	71.44	68.57	69.90
Peace	2,092	0.88	71.50	72.20	71.78
Force	1,107	0.71	66.94	67.54	67.22
All	2,596	0.93	80.71	74.42	77.39

Table 1: Tenfold cross-validation results from application of the naïve Bayes classifier for the different types of action coded for in the experiment. These results are also presented for merged action types. Each labeled row indicates the number of positive codings in the training data set, and the F_1-optimal posterior probability threshold for classification, in addition to its the corresponding values of precision, recall and combined F_1.

Action	No. pos. codings	Posterior prob.	Precision	Recall	F_1
Collective force	99	0.35	64.44	58.59	61.38
Collective peace	111	0.04	45.87	45.05	45.45
Singular force	6	0.42	0.25	16.67	20.00
Singular peace	96	0.11	44.68	43.75	44.21
Collective	168	0.17	75.91	61.90	68.20
Singular	101	0.36	41.18	55.45	47.26
Peace	178	0.23	53.69	61.24	57.22
Force	103	0.21	60.19	63.11	61.61
All	226	0.24	65.52	75.66	70.23

Table 2: Out-of-domain validation results from application of the naïve Bayes classifier for the different types of action coded for in the experiment. All data from Hong Kong are used for testing, and all other data are used for training.

References

Aggarwal, Charu C., and Karthik Subbian. 2012. "Event Detection in Social Streams." In *SDM*. SIAM / Omnipress pp. 624–635. http://dblp.uni-trier.de/db/conf/sdm/sdm2012.html#AggarwalS12.

Anastasopoulos, Lefteris Jason. 2015. "An Experiment on the Policy Effects of Immigrant Skin Tone." *Available at SSRN 2818629*.

Associated Press (AP) Image Database. 2016

Barberá, Pablo, Ning Wang, Richard Bonneau, John T Jost, Jonathan Nagler, Joshua Tucker, and Sandra González-Bailón. 2015. "The critical periphery in the growth of social protests." *PloS one* 10 (11): e0143611.

Bhavnani, Ravi, Karsten Donnay, Dan Miodownik, Maayan Mor, and Dirk Helbing. 2014. "Group segregation and urban violence." *American Journal of Political Science* 58 (1): 226–245.

Branton, Regina, Valerie Martinez-Ebers, Tony E Carey, and Tetsuya Matsubayashi. 2015. "Social protest and policy attitudes: The case of the 2006 immigrant rallies." *American Journal of Political Science* 59 (2): 390–402.

Carter, Gregg Lee. 1986. "The 1960s black riots revisited: city level explanations of their severity." *Sociological Inquiry* 56 (2): 210–228.

Clarke, James W, and Joseph Egan. 1972. "Social and political dimensions of campus protest activity." *The Journal of Politics* 34 (02): 500–523.

Collins, William J, and Robert A Margo. 2007. "The economic aftermath of the 1960s riots in American cities: Evidence from property values." *The Journal of Economic History* 67 (04): 849–883.

Dodds, Peter Sheridan, Eric M Clark, Suma Desu, Morgan R Frank, Andrew J Reagan, Jake Ryland Williams, Lewis Mitchell, Kameron Decker Harris, Isabel M Kloumann, James P Bagrow et al. 2015. "Human language reveals a universal positivity bias." *Proceedings of the National Academy of Sciences* 112 (8): 2389–2394.

Dodds, Peter Sheridan, Kameron Decker Harris, Isabel M. Kloumann, Catherine A. Bliss, and Christopher M. Danforth. 2011. "Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter." *PLoS ONE* 6 (12): 1-1.

Dong, Xiaowen, Dimitrios Mavroeidis, Francesco Calabrese, and Pascal Frossard. 2015. "Multiscale event detection in social media." *Data Min. Knowl. Discov.* 29 (5): 1374-1405.

Dunaway, Johanna, Regina P Branton, and Marisa A Abrajano. 2010. "Agenda setting, public opinion, and the issue of immigration reform." *Social Science Quarterly* 91 (2): 359–378.

Earl, Jennifer, Andrew Martin, John D McCarthy, and Sarah A Soule. 2004. "The use of newspaper data in the study of collective action." *Annual review of Sociology*: 65–80.

Epstein, Joshua M. 2002. "Modeling civil violence: An agent-based computational approach." *Proceedings of the National Academy of Sciences* 99 (suppl 3): 7243–7250.

Gerbaudo, Paolo. 2012. *Tweets and the streets: Social media and contemporary activism*. Pluto Press.

Huet-Vaughn, Emiliano. 2013. "Quiet Riot: The Causal Effect of Protest Violence." *Available at SSRN 2331520*.

Lipsky, Michael. 1968. "Protest as a political resource." *American Political Science Review* 62 (04): 1144–1158.

May, Paul. 2016. "Ideological justifications for restrictive immigration policies: An analysis of parliamentary discourses on immigration in France and Canada (2006–2013)." *French Politics*: 1–24.

Meirowitz, Adam, and Joshua A Tucker. 2013. "People Power or a One-Shot Deal? A Dynamic Model of Protest." *American Journal of Political Science* 57 (2): 478–490.

Metzger, Megan MacDuffee, Richard Bonneau, Jonathan Nagler, and Joshua A Tucker. 2016. "Tweeting identity? Ukrainian, Russian, and# Euromaidan." *Journal of Comparative Economics* 44 (1): 16–40.

Moore, Will H. 1998. "Repression and dissent: Substitution, context, and timing." *American Journal of Political Science*: 851–873.

Nurwidyantoro, A., and E. Winarko. 2013. "Event detection in social media: A survey." In *ICT for Smart Society (ICISS), 2013 International Conference on*. IEEE pp. 1–5. http://dx.doi.org/10.1109/ictss.2013.6588106.

Porta, Donatella della, and Sidney Tarrow. 1986. "Unwanted children: Political violence and the cycle of protest in Italy, 1966–1973." *European Journal of Political Research* 14 (5-6): 607–632.

Sayyadi, Hassan, Matthew Hurst, and Alexey Maykov. 2009. "Event detection and tracking in social streams." In *In Proceedings of the International Conference on Weblogs and Social Media (ICWSM 2009). AAAI*.

Seferiades, Seraphim, and Hank Johnston. 2012. *Violent protest, contentious politics, and the neoliberal state*. Ashgate Publishing, Ltd.

Sørensen, Rune Jørgen. 2016. "After the immigration shock: The causal effect of immigration on electoral preferences." *Electoral Studies* 44: 1–14.

Tarrow, Sidney. 1989. *Democracy and Disorder: Social Conflict, Political Protest and Democracy in Italy, 1965-1975*. New York: Oxford University Press.

Tarrow, Sidney G, and J Tollefson. 1994. "Power in movement: Social movements, collective action and politics.".

Tilly, Charles. 1978. *From mobilization to revolution*. Mcgraw-Hill College.

Tilly, Charles. 1988. *Collective violence in European perspective*. Center for Studies of Social Change, New School for Social Research.

Tilly, Charles, and Sidney G Tarrow. 2015. *Contentious politics*. Oxford University Press.

Torrens, Paul M, and Aaron W McDaniel. 2013. "Modeling geographic behavior in riotous crowds." *Annals of the Association of American Geographers* 103 (1): 20–46.

Tucker, Josh, Megan Metzger, Duncan Penfold-Brown, Richard Bonneau, John Jost, and Johnathan Nagler. 2014. "Protest in the Age of Social Media: Ukraines Euromaidan." *Carnegie Corporation*.

Wallace, Sophia J, Chris Zepeda-Millán, and Michael Jones-Correa. 2014. "Spatial and temporal proximity: examining the effects of protests on political attitudes." *American Journal of Political Science* 58 (2): 433–448.

Weidmann, Nils B, and Idean Salehyan. 2013. "Violence and ethnic segregation: a computational model applied to Baghdad." *International Studies Quarterly* 57 (1): 52–64.

Williams, J. R. 2016. "Boundary-based MWE segmentation with text partitioning." *CoRR* abs/1608.02025.

Williams, J. R., J. P. Bagrow, A. J. Reagan, S. E. Alajajian, C. M. Danforth, and P. S. Dodds. 2016. "Zipf's law is a consequence of coherent language production." *CoRR* abs/1601.07969.

Wilson, James Q. 1961. "The strategy of protest: problems of negro civic action." *Journal of Conflict Resolution*: 291–303.

Zhou, Xiangmin, and Lei Chen. 2014. "Event Detection over Twitter Social Media Streams." *The VLDB Journal* 23 (June): 381–400.

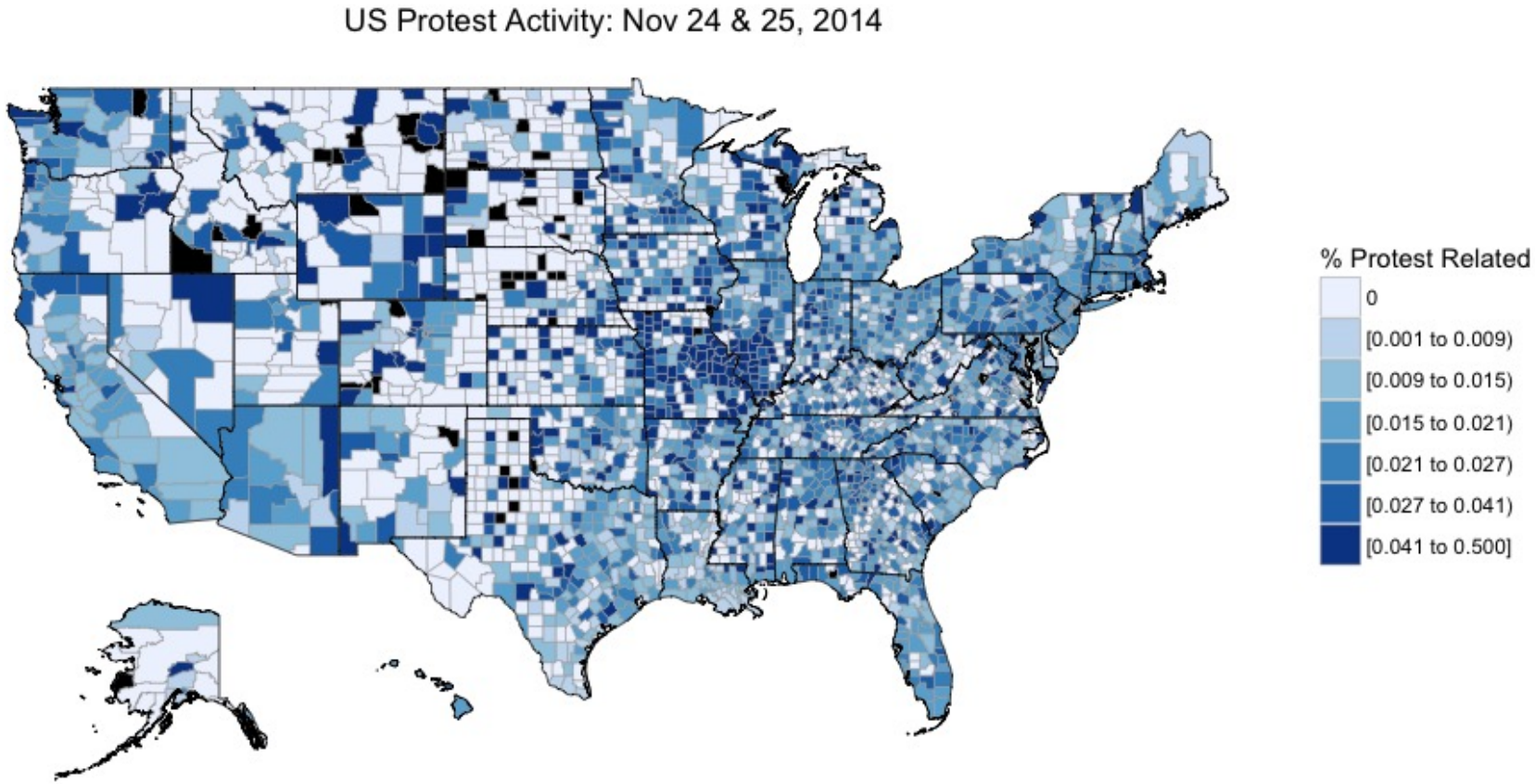

Figure 5: Ferguson Related Protest Activity Across US Counties on November 24th and 25th, 2014, Measure by % of Tweets Related to Collective Action

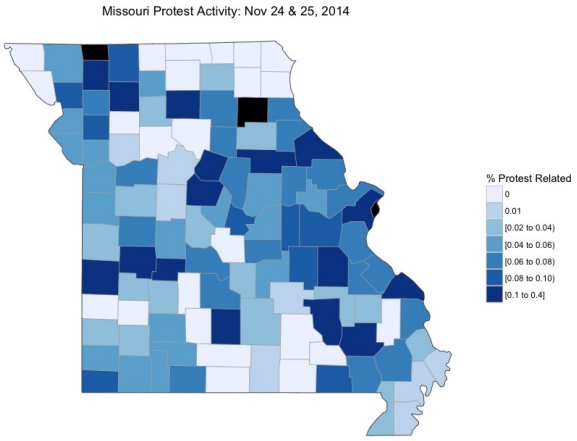

Figure 6: Ferguson Related Protest Activity in Missouri on November 24th and 25th, 2014, Measure by % of Tweets Related to Collective Action

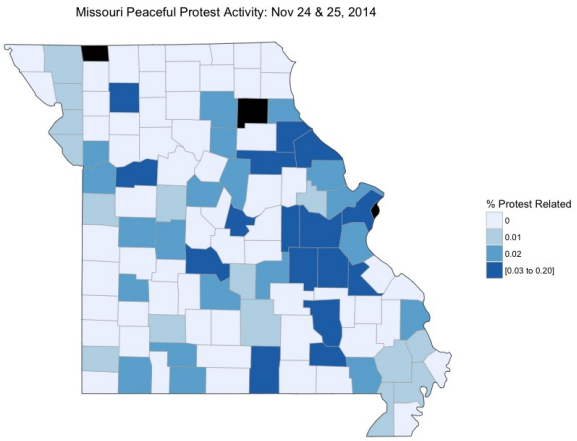

Figure 7: Ferguson Related Peaceful Protest Activity in Missouri on November 24th and 25th, 2014, Measure by % of Tweets Related to Collective Action

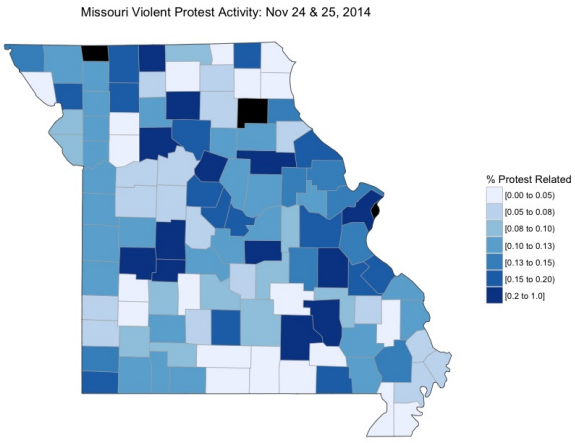

Figure 8: Ferguson Related Violent Protest Activity in Missouri on November 24th and 25th, 2014, Measure by % of Tweets Related to Collective Action