Social Movement Organizations in Online Protest Movements

ABSTRACT

In the era of Web 2.0, social media based Internet activism is slowly replacing traditional civic engagements such as sit-ins, strikes, rallies, etc. Proponents of online activism postulate that low participation and coordination costs, and the decentralized structure of these platforms will render the services and functionalities provided by Social Movement Organizations (SMOs) nonessential. Yet, such claims currently lack empirical evidence since many existing studies do not distinguish the participation of SMOs from individuals, and are therefore unable to determine the value of SMOs in online protest movements. This paper aims to fill that knowledge gap.

First, we use large-scale crowdsourcing platforms in combination with nested supervised learning models and identify more than 50 thousand SMOs participating in 2 distinct Twitter hashtag movements; next, by characterizing their primary objectives, we further assign each SMO to a specific social movement industry (SMI). We explore the attributes of these SMOs from five different perspectives: *commitment, knowledge sharing, community building, structural significance* and *recruitment*. We find that SMOs utilize a more diverse knowledge base, are more consistent and committed in their participation, apply more community-building efforts, possess more favorable positions within online protest networks, and contributed substantially to the recruitment of new participants compared to non-social movement organizations and individuals. We further compare and contrast behavior differences of SMOs from different SMIs. In summary, our findings show that the role of SMOs in social movements is far from over, even when focusing on online movements. In fact, social media affordances are allowing SMOs to participate together in a scope that is unattainable by traditional protesting methods.

ACM Reference format:
. 2018. Social Movement Organizations in Online Protest Movements. In *Proceedings of ACM Woodstock conference, Lyon, France, April 2018 (WWW'18),* 10 pages.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

At the peak of their influence, Social Movement Organizations (SMOs) such as the American Federation of Labor, held enough political sway to end child labor and pressure the Congress to pass the Fair Wage Act [30]. Not all SMOs are equally influential nor is any one SMO powerful perpetually. Nevertheless, SMOs have numerous advantages in comparison to isolated individual activists. SMOs generate social and intellectual capital [44]. SMOs are capable of keeping their members' ideologies alive in times of public apathy or hostility [56]. Members belonging to SMOs are more likely to participate in high risk activism than non-members [20, 39]. SMOs also provide the structure and leadership needed to incorporate a large population of new members from diverse background for swift and effective demonstrations [54].

Yet, in the age of social media, decentralized internet activism is slowly replacing traditional physically-engaged and often specialized advocacy methods. This has led some scholars to suggest that SMOs are less relevant in the era of online activism—reasoning that the decentralized nature of social media platforms favors more horizontal structures and that such platforms provide the affordances that drastically reduce organizing and participation costs, an issue often faced by traditional activists but mediated by the presence of SMOs in the past [18].

Research by Bailon et al. [24], however, rejects the supposed horizontal structure of online movements and suggest that the influence of individuals highly diverge and the underlying social networks of the participants are often sparse, containing many structural holes; thus, making it more difficult for online protests to cascade. Furthermore, analyses by Spiro et al. [53] indicate that while individual zealots are more likely to initialize protests online, SMOs eventually supersede any one individual activist as movements progress and become the center for broadcasting information and organizing movements.

Intrigued by the diverging observations, here we add to the current discourse by examining the role of SMOs within 2 online social media protests—the BlackLivesMatter movement and gender-related advocacies broadly categorized as Women's rights movement. We make the following contributions:

i.) To the best of our knowledge, we provide the first automated method to identify SMOs at scale. This method complements studies based purely on manual labor [46, 52]. As a result, we are able to identify a much larger and broader set of organizations.

ii.) We provide a richer understanding of SMO involvement in online protest movements compared to related work. In addition to analyzing the participation of SMOs in movements, we analyze their knowledge sharing and community building efforts, structural significance, and recruitment contributions.

iii.) We use automated unsupervised methods to identify the Social Movement Industry (SMI) corresponding to each SMO basing on its primary missions.

iv.) Using the SMI labels inferred from unsupervised methods, we provide a cross-industry analysis of SMO behavior, focusing on the BlackLivesMatter movement.

Our analyses reveals that SMOs still play an important role within social media movements. They are more active and committed to protests than non-SMOs and individuals; they draw significant fraction of recruits despite their relatively much smaller size, and they

WWW'18, April 2018, Lyon, France

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

are regarded by the broader social movement networks as more reputable than non-SMOs and individuals (demonstrated by their network position). Basing on these observations, we argue that SMOs are just as adapt, if not more so than individuals, at leveraging the affordances provided by social media and utilizing various communication strategies to enhance their online presence and attract potential recruits from a broad population.

Additionally, we see SMOs from an extensive set of varied industries including *African American, Youth, Student, Christianity, Research Institutes, LGBTQ, Social Welfare, Non-African American Minorities, Women&Health, Occupy Wall Street (OWS)* and *News*, protesting in solidarity or committedly in BlackLivesMatter, sustaining and expanding this movement with a joint effort. We observe a level of collaboration amongst SMOs that highlight the strength of social media for protest participation.

2 RELATED WORK

Our research draws from numerous classic theories and studies addressing SMOs within traditional participation media [20, 39, 42, 44, 56, 59, 60]. We adopt the the definition of social movement organization from the Resource Mobilization Theory proposed by McCarthy and Zald [42, 59]. In addition, we observe and adhere to the coding schemes published by McAdam [40].

Despite the significance of SMOs identified by studies listed above, most recent studies on online social movements fail to make a distinction between individuals and SMOs. Yet, there are a number exceptions that make that distinction [15, 26, 36, 43, 46, 47] that are worth noting here. Spiro et al. [53] examine the distinctions between zealots and SMOs in initializing and sustaining online social movements. They suggest that SMOs still play an important role in knowledge sharing and mobilization, especially in the mature stage of protests. Fetner and Key [21], by analyzing the role teapartypatriots.org played in the tea party movement, credit the movement's success to the website and the well financed and trained interest groups that built it. Conversely, research conducted by Lovejoy et al. [36] on social media usage by over 70 advocacy groups proposes that advocacy groups fell short of employing social media optimally. Furthermore, analyses done by Earl [18] suggest the waning importance of SMOs due to social media affordances.

This paper aims to contribute to the existing literature on the SMOs' social media strategies and success. Unlike the aforementioned studies, our study is based on a more comprehensive set of SMOs and is not limited by the most active organizations. This allows us to provide a more comprehensive understanding of social movement industries. Second, unlike related work that lists potential mechanisms to be used by SMOs, here we measure to what extent SMOs succeed in such actions. Finally, we measure SMO significance along multiple dimensions as opposed to focusing on a single one and study the differences between SMOs from different social movement industries (SMIs).

3 DATA

Our dataset consists of over 80 hashtag protests that are categorized into 2 broad topics (or, social movement industries): race-related activism collectively known as the BlackLivesMatter movement, and gender-related advocacies collectively grouped as Women's Rights movement. There are over 50 raced-related hashtag movements including #ferguson, #blacklivesmatter, #policebrutality, et cetera, spanning 36.6 million tweets and involving 4.3 million Twitter participants. Gender-related hashtags summate to 7.3 million tweets and 2.4 million advocates, covering issues including sexual violence, men's role in feminism, women's healthcare, women empowerment, wage equality, et cetera. The dataset includes *all* tweets with the corresponding hashtags between 02-2014 and 05-2015.

4 ACTIVIST CLASSIFICATION

Given Twitter does not supply account categories, we use crowdsourcing and supervised learning to classify the 7 million users into 3 distinct groups: SMOs, non-SMOs, and individuals.

4.1 Definitions

Social Movement Organizations (SMO): We define SMOs based on the Resource Mobilization Framework proposed by Zald and Ash, and McCarthy and Zald [41, 59]. Under this definition, an *SMO* "is an entity consisting an assembly of people that deliberately attempts to change individuals, established cultural norms, stati quo, institutions and structures, and/or to redistribute wealth." SMOs are distinct from non-SMOs in two ways. First, SMOs do not provide regular goods or services[1]. Second, SMOs' incentives are primarily purposive (deeply held principles or beliefs), secondarily solidary (social status, identity), and rarely monetary (wages, patronage).

As guided by the Resource Mobilization framework, in this work we categorize advancement and progress based non-profit, community, and student organizations, advocacy groups, interest groups, lobby groups, unions and employee associations, non-established political melioristic political groups, and proselytisic (conversion-seeking) religious entities[2] as SMOs [59]. Examples of SMOs and non-SMOs are shown in table 1.

Table 1: Organization Categorization and Examples

SMOs	
Subcategories	Examples
Advancement and Progress Based Non-profit, Community, and Student Organizations	Artist of Color; PRIDE Radio; Communities for a Better Environment in California; Graduate Student Employees Union
Advocacy Groups, Interest Groups, Lobby Groups	The American Civil Liberties Union; Ku Klux Klan; National Smokers Alliance
Labor Unions and Employee Associations	American Medical Association; United Automobile Workers
Proselytisic Religious Groups	Christianity; Islam
Non-established Political Organizations	Green Party; Socialist Party of America; the Tea Party
Non-SMOs	
Subcategories	Examples
Relief and Service Based Non-profit, Community, and Student Organizations	Student Accounts Services; Creative and Performing Arts; Social and Recreational Clubs; Neighborhood Watch; Doctors Without Borders
For-profit Organizations	Business and Finance; Corporate Media; Law Firms
Non-proselytisic Religious Groups	Buddhism; Modern era Judaism
Established Political Organizations, Societal Institutions and Structures	the Democratic Party; the Department of Labor; State Funded Public Educational Institutions

Social Movement Industries (SMI): Within the Resource Mobilization framework, an SMI is a group of SMOs that share broad

[1] Some organizations both provide services and engage in advocacy. We label these accounts by their primary objectives and functionalities, or as SMOs if primaries are unclear.

[2] All branches of Christianity, Islam, and Hinduism are classified as proselytisic religions, while modern era Judaism and Buddhism are considered nonproselytisic.

preferences or goals of a social movement [41]. Examples of SMIs include *Christianity* and *Environment Protection*[3].

Other Organizations (Non-SMO): Service-based nonprofit, community, and student organizations, established political entities, structures and systems, for-profit organizations (e.g. retailers, news organizations) and nonproselytisic religious groups are non-SMOs.

Individuals: Twitter accounts not classified as organizations will be coded as individuals.

4.2 Crowdsourcing Task

We recruit Amazon Mechanical Turk workers [11] to label a subset of Twitter accounts into one of the three aforementioned categories.

Sample Selection. A quick inspection of a random sample of Twitter accounts in our dataset reveals that SMOs (and organizations in general) are a rare class– there were 4 SMOs and 12 non-SMOs in a random sample of 200 users. To ensure a balanced sample for labeling, we stratify our dataset using 3 attributes organizations are likely to have: (1) having an URL listed under Twitter profile; (2) having more than 3 thousand followers[4]; and (3) being included on the homepage of any web domain listed under the *Society* and *News* categories on Alexa [1] or DMOZ [16][5]. We divide accounts into 6 strata: accounts that possess all 3 attributes, any 2 of the 3 attributes, any one of the 3 attribute, accounts that posted at least 2 tweets across all movements, accounts not included in previous strata, and finally, accounts selected completely at random [6]. We stratify and randomly select over 1600 accounts. This way, we are able to provide a wider coverage of the rare classes (SMOs and non-SMO organizations) while still sampling randomly from the broader distribution. In addition, we also obtain the dataset from Olteanu [47] which contain over 300 labeled organization accounts. Our final selected set has approximately 2000 accounts.

Instructions. We take screenshots of these accounts' Twitter homepages, randomly assign each screenshot into sets of 25, and export each set as an individual Mechanical Turk "human intelligence task" (HIT). Each HIT is assigned to three workers in order to assess label quality. Upon accepting a HIT, workers are first provided definitions of SMOs (listed in Section 4.1), examples (given in Table 1), and instructions for inclusions, exclusions, and exceptions. Next, they are taken through a training task where they are guided through the classification of 10 pre-labeled accounts. After the training task, they begin the actual labeling task [7].

Qualifications and Monitoring. Crowdsourcing platforms such as Amazon Mechanical Turk has successfully been used in a variety of labeling tasks in past studies (e.g. [10, 12]). However, this success relies heavily on the right mechanisms to ensure worker qualifications [38]. To ensure high-quality ratings, we required that workers: 1) reside in the U.S., (since the movements of interest were relevant to this geography) 2) had successfully completed at least 1,000 HITs; and 3) had an approval rate of at least 98%. Workers are first provided instructions about how to identify SMOs, then taken through a training process including 10 pre-labeled accounts prior to any actual labeling tasks. Furthermore, we allow workers to skip cases they are unsure about while still providing monetary compensation[8] and provide a feedback text box for additional user inputs. Such choices are shown to improve label quality [61]. Finally, we injected a number of Twitter accounts that were labeled by the researchers as a gold standard, monitored the accuracy of each worker for such cases and blocked workers if their accuracy fell below 60%.

Evaluation. Overall, despite the complexity of the task, the workers provided high quality and reliable labels with sufficient training and quality checks. We applied Krippendorff's alpha separately to assess inter-rater reliability. Our labels have a scores of 0.72, which is considered to have substantial agreement [51]. In addition, we also calculate Turkers' Gold set accuracy at 82%.

4.3 Supervised Classification Task

For supervised learning, we utilize a two-step classification process. We first build a binary classifier to label accounts into individuals and organizations, then for the accounts that are labeled as organizations, we apply a second classifier to further label them as an SMO or a non-SMO. Such a nested classification approach has been shown to work well in the case of imbalanced classes [10]. This approach also allows us to identify the optimal combinations of featuresets, preprocessing pipelines and classifier types, which yield better results than a regular multiclass classification method.

4.3.1 Featureset Selection.

Below, we list the set of features selected with the intent of identifying characteristics that discriminate individuals from organizations and/or SMOs from non-SMOs.

Username and user handle features We first derive sets of ngrams with 4-10 characters from (i.) username and (ii.) user handle fields for each account and extract two boolean features indicating whether they contain a person name using the python NLTK Name Corpus[9]. We also compute a boolean feature that indicates whether the user handle is included within any hyperlinks in the homepages of websites listed under Alexa and Dmoz's *News* and *Society* categories.[10]

Profile features We compute boolean features that signify whether account profiles contain (i.) an URL, (ii.) an occupation (using Kazemi's Occupation Corpus[32]), (iii.) numbers, (iv.) nonascii characters (e.g. emojis), and (v.) whether URLs, if present, belong to a social media website. In addition, we apply Flesch–Kincaid Grade to generate *readability*, an numeric feature indicating how difficult the profile text is to understand [34]. Furthermore, we apply Fasttext's Skipgram method, a shallow word embedding technique for text classification to associate each word token in account profiles to a numeric vector that represents the token's semantic dimensions [17, 31].

[3]For instance, *Greenpeace* and *Sierra Club* are SMOs both belonging to this SMI.
[4]This threshold is informed by related work on influential Twitter accounts [8]
[5]Alexa and DMOZ are website directories. Alexa is a forprofit company that hosts web traffic information on over 30M websites. DMOZ is an open directory project that has over 5M websites listed. We select *Society* given that its relevance to social movements (subcategories include *Activism*, *Ethnicity*, *Politics*, *LGBT*, etc.). In order to identify Twitter accounts listed under relevant subsections, we crawled the homepages of all the websites listed under *Society* and *News* for both sites, and gathered all the Twitter users contained within these homepages. Our crawler obtained over 500K website URLs and our dataset has 300K Twitter users listed under the homepages of these websites.
[6]A random sample is used as test set to reflect classifiers' performance against real-world distribution of accounts
[7]We will include links to such screenshots in the final version of the paper. We omit the links here in order not to jeopardize the double-blind process.
[8]In only a small fraction 3.7% of labeling tasks did the workers choose this option.
[9]The range 4-10 is based on surname length distribution published by the U.S. Census.
[10]We select *Society* given that it is most relevant to social movements (its subcategories include *Activism*, *Ethnicity*, *Politics*, *LGBT*, et cetera), and *News* because we expect media organizations to be the dominant type of non-SMOs engaged with protests.

We extract the word semantic vectors in the dimensions of 10, 25, and 50 to assess performance difference. Lastly, we use LIWC [49] to extract word dimensions such as *pronoun*, *work*, *social*, etc.

User motive-based features We generate features for total tweet count, protest related tweet count, protest related retweet count, and protest related mention count, all computed in log-scale. In addition, we calculate the ratio between users' total protest related tweets and users' overall number of tweets, the number of days between users' first and last protest tweets, number of distinct protest hashtags used, and the date of users' first protest tweet. This set of features represents users' overall involvement and intensity in protests.

User influence-based features Influence-based features include: the number of friends and followers (both in log-scale), the ratio between the two, and the number of retweets and mentions users received from others (both in log-scale). This featureset represents users' influence or prestige on Twitter.

Protest Tweet Features This set includes the number of original tweets (i.e. not a retweet), the number of urls, the number of hashtags, and the number of non-ascii characters used in protest related tweets. In addition, similar to user profile text, we again apply Flesch–Kincaid Grade to obtain a continuous measure of tweet readability, LIWC to extract word dimensions, and Fasttext's Skipgram method to generate word to vector features for the original tweets.

4.3.2 Classifier Selection. In total, 7 classifiers are evaluated for each classification task. Details are provided below.

Single Classifier We first select 5 commonly used classifiers: Linear Regression[33], Linear Support Vector Machine (SVM)[19][11], Random Forest[7], Extra Trees[23], and K-Nearest Neighbor[14] as implemented by Python's *Scikit-learn* library[48]. to compute baseline performance.

Majority Voting Classifier We generate $5C3$ (5 choose 3) combinations of classifiers from the single classifier list and assign the selected 3 as base estimators to a majority-voting based ensembler as implemented by *Scikit-learn*.

Stacking Classifier We apply another ensemble classification technique called *Stacking* [55, 58] from the *MLxtend* library. This method combines multiple classification models via a 2nd level meta-classifier where the prediction outputs by the level-1 classifiers are used as inputs to the meta-classifier. Again, we generate $5C3$ combinations of single classifiers as the level-1 classifiers, each remaining classifier not in level-1 is used as the level-2 classifier, or the meta classifier.

4.3.3 Classification Performance. For each classification pipeline, we apply featureset selection, preprocessing, and hyper-parameter tuning to derive the best fit. We assess classification performance based on (1) testset including datapoints from all the strata and (2) testset only including the random stratum. For the former, we apply a stratified ten-fold cross validation; for the latter, all non-random strata accounts are assigned to trainingset, we then utilize a three-fold cross validation on the random strata accounts to estimate performance. We measure the best combinations of featuresets, preprocessing steps, and classifier selections using the sum of the rarer class's precision and recall.

[11]Linear SVM by default does not provide probability based decision function, we apply Platt's scaling to derive the prediction probability [50].

Individual and Organization Classification. Detailed evaluation statistics for the best classifiers including accuracy, precision, recall, and AUC scores estimated are summarized in Table 2. Due to space limitations, here we only present results for the stacking, majority voting and the best performing single classifier (Extra Trees classifier for this task). All our classifiers for both test-set types have AUC >0.9 which is considered *excellent* accuracy [6]. Our best classifier is the stacking classifier with Linear SVM, Random Forest, and Extra Trees as level-1 classifiers, and Linear Regression as the meta classifier which has an accuracy of 0.97 and an AUC score of 0.98. Using this classifier, we label 312K accounts in our dataset as organizations. We also inspected the features with the highest weights to gain insights about the classifier. Not surprisingly, having an actual person's name, including first person pronouns or an occupation in the profiles strongly indicate accounts as individuals.

SMO and non-SMO Classification. For our second classification step, we only use accounts classified as organizations from the previous classification step. Table 3 summarizes our best classifiers' performance. Due to space limitations, here we only present results for the stacking, majority voting and the best performing single classifier (Linear SVC for all strata and Random Forest for the random stratum). When testing against all strata, our classifier achieves an accuracy of 0.87 and an AUC score of 0.83; when testing against random stratum, our classifier has an AUC score of 0.77 with SMO precision and recall are both approximately 0.7. Furthermore, an inspection of feature weights reveal that the ratio between total protest tweets and all tweets posted by accounts, and being retweeted or mentioned more by other participants suggest accounts as SMOs instead of non-SMOs. We apply the best performing classifier to the set of 312K accounts classified as organization by the individual/org classifier and identify more 50 thousand accounts as SMOs.

To ensure a more accurate assessment of SMOs' role in social media activism, we also examined the most active [12], or most influential [13] participants within our dataset. We found over 700 highest value participants within our labeled organizations dataset and manually reassigned them into correct categories as SMOs, non-SMOs, or individuals. Accuracy measures for this sample was comparable to the findings in the test set. For instance, approximately 78% of SMOs were correctly identified by our classifier.

Table 2: Individual/Organization Classification

strata	type	accuracy	precision indv	precision org	recall indv	recall org	auc
all	stack	0.97	0.96	0.98	0.99	0.93	0.98
all	voting	0.92	0.91	0.94	0.97	0.83	0.95
all	single	0.91	0.91	0.93	0.96	0.84	0.94
random	stack	1.00	1.00	1.00	1.00	1.00	1.00
random	voting	0.99	0.99	0.95	1.00	0.90	1.00
random	single	1.00	1.00	1.00	1.00	1.00	1.00

5 ANALYSIS

We first identify and remove the subset of SMOs that belong to the countermovements of BlackLivesMatter and Women's Rights

[12]>=99.99 percentile in number of protest tweets contributed.

[13]>=99.99 percentile in both number of retweets and mentions received.

Table 3: SMO/non-SMO Classification

strata	type	accuracy	precision nonsmo	precision smo	recall nonsmo	recall smo	auc
all	stack	0.87	0.88	0.83	0.92	0.76	0.83
all	voting	0.86	0.86	0.86	0.94	0.71	0.82
all	single	0.84	0.84	0.85	0.94	0.66	0.84
random	stack	0.88	0.93	0.68	0.92	0.69	0.77
random	voting	0.89	0.91	0.79	0.96	0.56	0.77
random	single	0.88	0.91	0.75	0.96	0.54	0.77

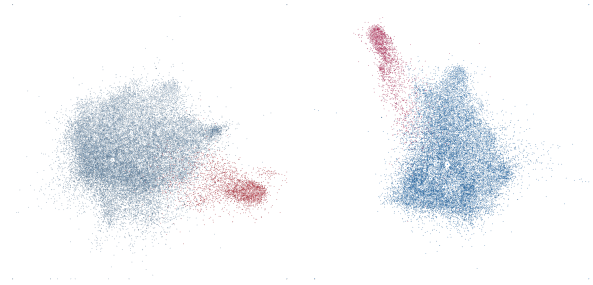
in Section 5.1. We then examine the importance of SMOs in comparison to individuals and non-SMOs for each movement in Section 5.2. Next, in Section 5.3, we focus on the BlackLivesMatter movement and compare organizations from different SMIs. For both analysis, we focus on five dimensions: *commitment, knowledge sharing, community building, structural significance* and *recruitment*. We select these dimensions basing on the traditional functions of SMOs as well as the recent researches studying SMOs' social media strategies[2, 26, 46, 59].

5.1 Countermovement SMOs

Previous researches focusing on Twitter hashtag protests identified numerous strategies widely employed by counter-protests. Organizations and individuals who fundamentally disagree with a protest's narrative and goals would utilize strategies such as alternative narrative or framing[9, 45], hashtag hijacking and spaming[13, 27, 29], establishing and spreading counter-protest hashtags [22], et cetera, to attack and discredit the "offending" movement. To ensure the SMOs under examination are in fact supportive of BlackLivesMatter and Women's Rights, we first apply network clustering to identify and remove the subset of SMOs that belong to the countermovements.

First, we define two directed graphs $G_{blm} = (N_{blm}, E_{blm})$, and $G_{women} = (N_{women}, E_{women})$. Here a node $u \in N_{blm}$ if u is an SMO and u participated in the BlackLivesMatter movement by using at least one of the related hashtags at least once. A directed edge $e_{u,v} \in E_{blm}$ from u to v if u is followed by v. We apply the same procedure to generate G_{women}. We then use Gephi[3], to visualize G_{blm} and G_{women} [14]. As shown in figure 1, both BlackLivesMatter and Women's Rights contain a small cluster of nodes, approximately 10% in size and colored in red, that are far removed from the main cluster. To ascertain that these are in fact countermovement SMOs, we adopt from related research by Starbird[45] which shows that countermovement Twitter accounts strategically include hashtags in the profile descriptions field to enable account owners to easily find each other and establish a network for amplification purposes. We examine the BlackLivesMatter SMOs colored in red and discover that the top 5 most frequent hashtags in their profiles are *tcot, teaparty, tgdn, freedom*, and *2a*. In addition, we randomly sampled 100 tweets, each posted by a distinct account from the red subset, and manually label them as supportive of the BlackLivesMatter movement or opposing. We note that 79 tweets objected the movement and framed the protesters as disillusioned or mobs. We conduct similar analysis on Women's Rights movement and observe a comparable pattern. We remove these accounts for the subsequent analyses.

[14]Gephi is an open source software for network exploration and manipulation. We use its build-in graph layout and network clustering features: *ForceAltas2*, and *Modularity*[4]



(a) BlackLivesMatter (b) Women's Rights

Figure 1: SMOs Follower-followee Network Clustering.

5.2 Analyzing SMOs, non-SMOs, and Individuals

5.2.1 Direct Contribution. There were 33.7K and 19.6K SMOs involved in BlackLivesMatter and Women's Rights respectively. Comparing that to the Civil Rights movement of the past where SMOs participated in the range of hundreds [62], it is evident that same affordances that lower costs for participation for individuals are leveraged by SMOs as well, despite the claims of related work that SMO relevance is reduced due to social media affordances[18]. The significance of SMOs are further demonstrated when considering other participation measures listed in Table 4. For instance, for BlackLivesMatter, while SMOs only account for 0.7% of all participants, they contributed 2.2 million tweets amounting to 6.0% of all tweets. Furthermore, BlackLivesMatter SMOs received 1.4 million mentions (14.6%), and 4.2 million retweets (15.4%). On average, a BlackLivesMatter SMO contributes 64 tweets, and receives 124 retweets and 40 mentions—significantly greater than the corresponding measures for non-SMOs and individuals.

5.2.2 Commitment. We measure commitment in two ways. First, we calculate the participation lifetime of protesters by computing the number of days elapsed between the first and last tweet in a given movement. Second, we compute the fraction of SMOs, non-SMOs and individual accounts actively participated in the corresponding protest per day averaged over time. As shown under the commitment section in Table 4, for BlackLivesMatter protest, an average SMO has a lifetime of 82 days, while an non-SMO or an individual has a lifetime for 69, and 45 days on average. Likewise, 2.3% of all BlackLivesMatter SMOs are actively participating in the protest per day averaged over time in comparison to 1.2% of all non-SMOs, and 0.6% of all individuals. This pattern is also observed for Women's Rights. Furthermore, we find that SMOs indeed show higher commitment even when accounting for the number of tweets they generated[15].

5.2.3 Knowledge Sharing. Previous studies on advocacy organizations label these organizations' social media strategies into 3

[15]We fit the regression model: $D_{blm,i} = \beta_0 + \beta_1 B_{blm,i} + \beta_2 C_{blm,i} + \varepsilon_i$ where $D_{blm,i}$ is the number of unique days i has participated in the BlackLivesMatter protest, $B_{blm,i}$ is number of tweets by i in the protest and $C_{blm,i}$ is the category of user i (SMO, non-SMO or individual). $G_{blm,i}$ is significant and large for SMOs. A similar pattern is also observed for Women's Rights.

Table 4: SMOs Contribution Overview

Dimension	Measurements	BLM SMO	BLM Non-SMO	BLM Indv	Women SMO	Women Non-SMO	Women Indv
Direct Contribution	% accounts	0.7%	3.6%	95.7%	0.8%	3.3%	96.0%
	% tweets	6.0%	6.8%	87.2%	3.0%	4.4%	92.6%
	% mentions received	14.6%	24.0%	61.4%	25.8%	9.5%	64.6%
	% retweets received	15.4%	10.2%	74.2%	11.0%	7.0%	82.0%
	mean/median tweets	64/2	15/2	7/1	11/2	4/1	3/1
	mean/median mentions of	40/0	14/0	1/0	34/0	3/0	1/0
	mean/median retweets of	124/0	17/0	5/0	28/1	4/1	2/1
Commitment	mean/median protest length in days	82/1	69/1	45/1	78/1	43/1	29/1
	mean/ median fraction of accounts active per day	2.3%/0.7%	1.2%/0.3%	0.6%/0.1%	1%/0.7%	0.5%/0.2%	0.3%/0.1%
Knowledge Sharing	unique URLs	177K	187K	615K	37K	48K	211K
	unique domains	9.9K	14.6K	29.0K	5.8K	9.5K	21K
	mean/median URLs	14/1	3/0	1/0	3/1	1/1	0/0
Community Building	mean/median unique protest hashtags	4/2	3/1	2/1	2/1	1/1	1/1
	mean/median protest communities	2/1	2/1	1/1	1/1	1/1	1/1
Structural Significance	mean/median kcore	31/4	13/3	6/2	7/3	4/2	3/2
	mean/median indegree-centrality	3.1e-5/9.2e-7	9.0e-6/6.9e-7	2.3e-6/4.6e-7	2.5e-5/1.3e-6	5.1e-6/8.9e-7	2.2e-6/8.9e-7
Recruitment	total recruits	356K	586K	2.4M	467K	167K	1.9M
	% recruits	10.7%	17.6%	71.6%	18.7%	6.7%	74.6%
	mean/median recruits	11/0	4/0	1/0	22/0	2/0	1/0

broad categories: knowledge sharing, community building, and mobilization [2, 26, 46], and identify knowledge-sharing as the primary strategy. Their more frequent use of URLs provide additional content to their network, unhinged by Twitter's 140 character limit. Informed by these studies, we measure knowledge sharing based on the number of hyperlinks shared. As shown in Table 4, BlackLivesMatter SMOs shared a total of 177K unique links, almost equating to the amount shared by non-SMO organizations despite being only $\frac{1}{4}$ the size. Furthermore, an average SMO shares 14 links, while a non-SMO only shares 3, and an individual a single link. Similar, though less remarkable, results are obtained for Women's Rights. The higher levels of knowledge sharing can, much like commitment, be due to the higher rate of tweeting. Similar to Section 5.2.2, we investigate and find that SMOs engage in knowledge sharing beyond what is expected simply due to their activity levels [16].

5.2.4 Community Building. Research based on a limited set of SMOs suggests that SMOs use hashtags more frequently than the baseline usage on Twitter [26]. This is considered a community-building effort and a mobilization strategy utilized by the advocacy groups to make their tweets available for hashtag based search on Twitter. Building on this research, we first examine whether SMOs are active in community building by using more movement related hashtags (e.g. #ferguson and #blacklivesmattter as opposed to only #blacklivesmatter). Next, we go beyond simply reporting the number of hashtags used by activists and instead examine which communities the activists reach out to through their hashtags. To achieve this goal, we construct hashtag-hashtag co-occurrence graphs and cluster similar hashtags into communities. We defined similarity between hashtags $\#h_i$ and $\#h_j$ as the Jaccard similarity between T_i and T_j where T_i (T_j) is the set of tweets containing hashtags $\#h_i$ ($\#h_j$). We then build a graph where an edge exists between pairs of hashtags that have non-zero Jaccard similarity and the Jaccard similarity score denotes the edge weight. With the constructed graph, we apply Louvain heuristics to compute partitions of highest modularity [4] and have identified 7 communities for BlackLivesMatter and 5 communities for Women's Rights[17]. The results are visualized in Figure 2 and summarized in Table 4. On average, BlackLivesMatter SMOs use 4 different protest hashtags, while non-SMOs use 3, and individuals 2. SMOs indeed use more hashtags even when accounting for participation intensity levels [18]. Communities-wise however, majority participants chose to focus on a single community.

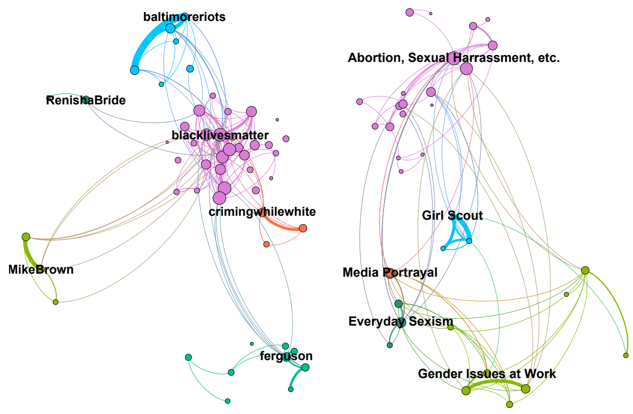
5.2.5 Structural Significance. We use two measures: 1) *in-degree*, which measures the number of ties directed at a participant and captures a participant's popularity or reputation [5], and 2) *k-core* score, which was used by [25] to measure a participant's ability to seed network cascades in analyzing protest recruitment through online networks. We measure both values to assess SMO structural significance. First, we define two directed graphs $G_{blm} = (N_{blm}, E_{blm})$, and $G_{women} = (N_{women}, E_{women})$. Here a node $u \in N_{blm}$ if u participated in the BlackLivesMatter movement by using at least one of the related hashtags at least once, and an edge $e_{u,v} \in E_{blm}$ if u mentions or retweets v at least once (similar construction for G_{women}).[19] We compute the in-degree and k-core value of nodes corresponding to SMOs, non-SMOs, and individuals. The mean and median of

[16] We fit the regression model $D_{blm,i} = \beta_0 + \beta_1 B_{blm,i} + \beta_2 C_{blm,i} + \varepsilon_i$, where $D_{blm,i}$ is the number of unique hyperlink domains shared by i, $B_{blm,i}$ is number of tweets by i in the protest and $C_{blm,i}$ is the category of user i (SMO, non-SMO or individual). The results show that being an SMO is positively correlated with an account sharing from more diverse domains.

[17] Sample Hashtags belonging to the same community are #wearesilent and #wearehere. Due to space limitations, we cannot list all communities.

[18] We fit the regression model $H_{blm,i} = \beta_0 + \beta_1 B_{blm,i} + \beta_2 C_{blm,i} + \varepsilon_i$, where $H_{blm,i}$ is the number of unique protest hashtags by i, $B_{blm,i}$ is number of tweets by i in the protest and $C_{blm,i}$ is the category of user i (SMO, non-SMO or individual). The results show that being an SMO is positively correlated with an account using more unique hashtags.

[19] We also analyzed structural patterns when edges $e_{u,v}$ are based solely on retweeting or solely on mentioning behavior. The results were consistent with what is presented here.



(a) BlackLivesMatter (b) Women's Rights

Figure 2: Hashtag Movement Communities. For instance, the community marked as *baltimoreriots* contains hashtags such as #baltimoreuprising and #baltimoreriots that are closely linked to the incidents at Baltimore. In Women's Rights movement, the community of hashtags labeled *Gender Issues at Work* includes #changetheratio, #ask4more. Edges with low weights are removed for ease of viewing.

these distributions in Table 4. SMOs have higher average in-degree and k-core values, which would imply that they are popular, and their network position grant them a more favorable probability of initializing cascades of online movements.

5.2.6 Recruitment. Lastly, we evaluate the success of SMOs strategies by determining the number of activists SMOs recruited into movements on Twitter. We measure the number of recruited activists as follows: We first identify the set of activists A_{blm} and A_{women} that participated in BlackLivesMatter and Women's Rights movements by using at least one relevant hashtag. Next for each activist $a_{blm} \in A_{blm}$ (and similarly for Women's Rights), we identify the first tweet $t_{a_{blm}}$ of a_{blm} in that movement. We identify whether $t_{a_{blm}}$ mentions or retweets another activist a'_{blm}. If so, we denote a' as the recruiter of a into the corresponding movement. Then we determine whether a'_{blm} is an SMO, a non-SMO, or an individual. If the first tweet by a_{blm} contains multiple mentions and/or retweets, each mentioned/retweeted account is considered a recruiter. The results are summarized in table 4. We observe that the average number of recruits by BlackLivesMatter SMOs is 11, and non-SMOs is 4 and only 1 for individuals. SMOs are in general the most active group in terms of recruitment per account. This pattern is consistent for Women's Rights.

5.3 Comparing SMOs from Different SMIs

Thus far, we addressed protest behavior differences between SMOs, non-SMOs, and individuals–treating all SMOs as part of a single group. Yet primary objectives amongst SMOs vary widely. Indeed, within the Resource Mobilization framework, McCarthy and Zald acknowledge this and define Social Movement Industry (SMI) as the group of SMOs sharing broad preferences or goals of a social movement [41]. Traditionally, SMOs largely focus on causes related to their SMIs. Yet, the new era of online activism stands to change this pattern. For instance, an SMO with the primary goal of empowering young girls can easily show solidarity to a social movement of another cause, say BlackLivesMatter, while spending the majority of its capital and efforts on Women's Rights movement. Consequently, the behavior of organizations from different SMIs can vary drastically. Here, we first use unsupervised methods to cluster SMOs into different SMIs; we then analyze and compare the behavior of SMOs in different SMIs. Due to space limitations, we only focus on the BlackLivesMatter movement.

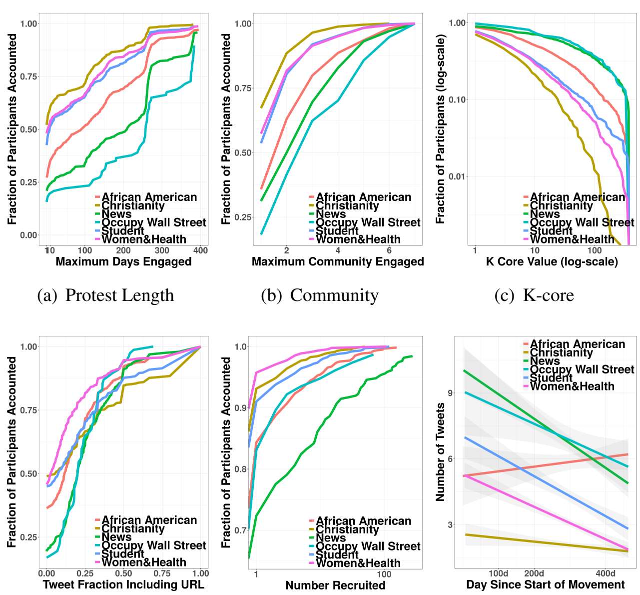
5.3.1 Identifying SMI for each SMO. We use k-means clustering of SMO profile descriptions, coupled with a rigorous preprocessing step, to categorize SMOs into SMIs.[20] For each SMO profile description, we first remove nonascii characters and repetitive punctuations, and assign part-of-speech (POS) labels to each token.[21] We focus on noun tokens given their strength in extracting semantically meaningful topics [37]. For the set of nouns derived, we examine the subset that occur 100 times or more to assess their semantic value, and remove the ones that lack contextual cues.[22] In addition, we manually examine the most frequently occurring non-noun tokens and selectively keep the ones that are contextually relevant.[23] This results in a dictionary of size 1.6K. We keep the set of profiles that have 4 or more relevant tokens and discard the rest, resulting in 22K matching accounts. For the remaining accounts, we first translate tokens into a sparse document term matrix and apply the elbow method [57] to determine the optimal number of document clusters, which is 24, and then use kmeans clustering.

To assess the unsupervised method's ability to correctly assign SMOs to SMIs, we randomly select a sample of 25 accounts from each cluster and then manually label each profile as correctly clustered or not. Using this method, we identity 12 distinct SMIs with random sample accuracy of 70% or above: *African American, Youth, Student, Christianity, Research Institutes, LGBTQ, Social Welfare, Non-African American Minorities, Women&Health, Occupy Wall Street (OWS)* and *News*. [24] To inspect the differences between the 12 SMIs, we reapply our analyses from the previous section including *commitment, knowledge sharing, community building, structural significance* and *recruitment*. Due to space limitations, we only select and discuss a representative subset of the SMIs: *African American, Christianity, News Media, Student, Women&Health* and finally, *OWS*. Three randomly selected account profiles, their corresponding tokens, and clustering results for each SMI in the subset are listed in Table 5. In the next section, we present results for a subset of the measures for the 5 dimensions. The omitted results are in line with the analysis listed here.

[20] We explored the use of LDA (latent dirichlet allocation) and have found the topics to be less coherent. We also attempted to use profile similarity measurements between pairs of SMO accounts to create weighted edges and apply existing community detection algorithms on the corresponding network, which was also unfruitful.
[21] We apply *spaCy*, a natural language processing library that implements parsing and tagging algorithms developed by [28, 35] to assign POS labels.
[22] For instance, the word "community", does not provide as much context as "violence".
[23] Sample non-noun tokens kept include"black"—an adjective, but highly relevant to the context; "occupy", normally a verb, but here signals relation to the Occupy movement.
[24] We label the SMI clusters identified using the coding schemes by McAdam[40]. Note that *News* SMOs differ from mainstream corporate media firms in that *News* SMOs center their organizational objectives around shining light into social, political, or economic issues, and/or bringing about positive depictions of certain groups.

Table 5: SMI Clustering, Sample Profiles, and Generated Tokens

SMI Cluster	Account Profile	Generated Tokens
Christianity	*1.these are updates, information and prayers for the #downtown congregation of #redeemer presbyterian #church.*	update, information, prayer, congregation, presbyterian_church
	2.north coast training helps you craft strong, stable ministries that prepare your church to break though barriers, overcome plateaus, and prime for growth.	coast, training, stable, ministry, church, barrier, growth
	3.hello from westside unitarian universalist church! we're proud to be a progressive, welcoming voice in ft worth. check out our website & drop in for a visit.	unitarian, universalist, church, proud, progressive, voice, worth, website, visit
News	*4.a new kind of social network, democratic news for the people, by the people. #n#join the revolution...*	kind, network, democratic, news, revolution
	5.christian retailing serves the christian products industry worldwide. visit our website to sign up for our print/digital issue or e-news or to download our app.	christian, christian, product, industry, website, print, digital, news, app
	6.leaders, solutions, and news from the global majority found around the world!	leaders, solution, news, majority
Occupy Wall Street (OWS)	*7.catholic $ #us $ green party occupy-comrade ready 4 merry marriage 2 create souls :) we profit 4 god/jesus: human soul-u-tion; world #lovevolution! 323.258-4434*	catholic, green, party, occupy, marriage, soul, god, jesus, soul
	8.occupy, politics, anti-corruption, anti-megacorp, against dangerous levels of power in any form. for the people.	occupy, politic, corruption, form
	9.west hollywood division of occupy los angeles. be on the right side of history. stands in solidarity with #ows #ola #oo #hrc.	west, division, occupy, angele, side, solidarity
African American (AA)	*10.black is beautiful#n#black lives matter#n#civil rights for all#n#love unconditionally*	black, black, matter, civil_right, love
	11.#blacklivesmatter is an affirmation and embrace of the resistance and resilience of black people.	blacklivesmatter, affirmation, embrace, resistance, resilience, black
	12.an auxiliary organization existing on college campuses nationwide directly related to, and in turn shares the ideas and views of, the 100 black men of america.	college, campus, turn, view, black
Student	*13.fair employment project is a non-profit legal-aid organization of attorneys, students, and worker advocates dedicated to employment civil rights.*	legal, aid, attorney, students, worker, employment, civil_right
	14.students against destructive decisions. we are a club and we encourage teens to make responsible life choices.	students, decision, club, teen, choice
	15.the mission of the office of student development and orientation is to educate, enhance, and enrich students and their collegiate experience.	office, orientation, students, collegiate, experience
Women&Health	*16.a center for women and their families living with hiv/aids and other health disparities including hepatitis c, obesity, diabetes, hypertension and more.*	women, family, hiv_aid, health, disparity
	17.we offer a safe environment for all women to come together to gain confidence, develop skills, get support and much more!	safe, environment, women, confidence, skill
	18.founded in 2009 as a #feminist & young women-led org working for women's sexual & reproductive health rights (srhr), leadership & dev #cameroon.	young, org, women, reproductive, health, leadership, dev



(a) Protest Length (b) Community (c) K-core

(d) URL Usage (e) Recruited by Mention (f) Participation Intensity

Figure 3: SMO Protest Attribute Differences Given SMI Context.

5.3.2 SMO Comparisons With SMI Context. As shown in Figure 3, SMO accounts belonging to the *African American* SMI, denoted as A_{aa}, are more central to BlackLivesMatter than the subset of SMOs that participated solidarily such as $A_{christianity}$ and $A_{women\&health}$. The median tweet contribution by A_{aa} is 8, and 2 for both $A_{christianity}$ and $A_{women\&health}$; in addition, approximately 50% of all A_{aa} participated more than 100 days in BlackLivesMatter, while only 25% of all $A_{christianity}$ did the same. A_{aa} are also more active in community building with 22% of all A_{aa} engaged in 3 or more protest communities. In comparison, less than 10% of $A_{christianity}$, $A_{women\&health}$, and $A_{student}$ depicted similar effort. Similar patterns are also observed in knowledge sharing, structural significance and recruitment rate.

Interestingly, both A_{news} and A_{ows} exceed A_{aa} in all 5 dimensions. For instance, the median tweet contribution by A_{news} is 32, 4 times that of A_{aa}. The median number of retweets and mentions received by A_{news} are 20 and 4 respectively while the same numbers for A_{aa} are 2 and 1. Furthermore, A_{news} is also the most successful at recruiting new participants. The recruitment rate for A_{news} is 0.83 per tweet, approximately 4 times that of the second highest group A_{ows}. Additionally, the median number of URLs contributed by A_{news} is 7 in comparison to A_{aa}'s single URL. A_{news} also shared information from 14K unique URLs comparing to the 7K, and 4K shared by A_{aa} and A_{ows}. These observations suggest that A_{news} is much better at employing social media mechanisms and participating in *knowledge sharing* and *recruitment* perhaps due to the nature of its industry.

While the critical role of A_{news} in BlackLivesMatter can be explained by the need of engaging others and spreading awareness, what's more surprising is that A_{ows} surpasses both A_{news} and A_{aa} in their *commitment* to BlackLivesMatter, despite it being outside of their primary SMI. The median tweet contribution by A_{ows} is 54, the highest amongst all the groups. In fact, as depicted in Figure 3, A_{ows} is the most dedicated amongst all SMO groups as measured by protest length and number of protest communities engaged. Consequently, A_{ows} also possess the highest structural prestige. The median kcore value for A_{ows} is 55 comparing to A_{news}'s 41 and A_{aa}'s 11.

The firm efforts by and structural centrality of A_{ows} were initially surprising. Further analysis revealed a plausible explanation—our dataset capture the early stage of *African American* SMOs' engagement in BlackLivesMatter where movement specific SMOs have yet to fully mature.[25] Indeed, over 35% of all A_{aa} accounts on Twitter were created *after* February, 2014, a time when the BlackLivesMatter movement started to gain national recognition. This number is only 2% for A_{ows}. The Occupy movement, which started in 2011, was well developed with a mature and stable social media presence by 2014. Furthermore, as the BlackLivesMatter movement progresses, the protest activity level, measured by tweet count, of A_{ows} wanes when that of A_{aa} rises as shown in Figure 3(f)[26]. We see that A_{aa} eventually supersedes A_{ows} in protest intensity. In fact, A_{aa} is the only SMI that depicts a rise in participation level. This is consistent with Resource Mobilization[42] which posits that as a movement grows and gains capital (human, social, or monetary), professional movement-specific SMOs will eventually emerge.

6 DISCUSSION

In this paper, we set out to characterize the importance of Social Movement Organizations for online political movements using two significant social media protests: race related hashtag movements generally referred to as BlackLivesMatter and gender-related hashtag movements broadly categorized as Women's Rights.

The study focused on five dimensions of SMO significance: *commitment, knowledge sharing, community building, structural significance* and *recruitment*. All five analyses point to the same overall finding: the claim that SMOs importance is waning is largely incorrect. Social media affordances provided to individuals are just as present to SMOs. Furthermore, considering SMOs possess financial, human, and social capitals that are less readily available to isolated individuals, and that addressing societal issues being their prime directive, we see that SMOs to be far more committed to the movements than individuals and non-SMOs; that they are better at leveraging social media utilities for their cause, and play a critical role in knowledge sharing and community building efforts. Consequently, we also observe SMOs to possess greater structural significance (or, network prestige) within the online protest communities, and more successful at recruiting new participants.

Yet, it is unrealistic to assume there are no shifts. Indeed, we observe that certain subset of individuals do hold strong positions in movements, with some being able to recruit more protesters that the most active SMOs. In addition, we also see that a large number of SMOs from varied SMIs participated in solidarity, lending their Twitter network to assist a social movement not of their immediate domain (as exemplified by the surprisingly assertive role of Occupy SMOs within the BlackLivesMatter movement). We posit that social media affordances are carving out new pathways for SMOs to engage and cooperate with each other, allowing them to easily and cheaply pool in certain types of social capital to sustain and expand online social movements. It will be crucial for future work to unpack these new dynamics, compare behavioral differences, cooperations (or even competitions) amongst the different groups, across varied movements, times, and locations.

This study peels the first layer and finds that SMOs are adapting to the new era of online activism by engaging heavily in social media efforts. It is also worthwhile to note various significant challenges addressed in this paper. The first challenge was to develop a crowdsourcing task with the right set of coding instructions and worker qualification schemes to reliably label Twitter accounts as SMO, non-SMO, or individuals. The research team has gone through a number of iterations to create a task with high inter-rater reliability and accuracy measures. Traditional efforts in identifying SMOs rely on experts for coding and have only focused on news media content [40]. Here, we demonstrate the ability of crowd workers to provide reliable coding with a short training session and we extend the coding mechanism from one that uses long news articles to one that uses social media data. The movements studied in this paper were high-profile, which might have made the labeling task easier. Whether the crowd can be relied on to detect SMOs across different SMIs is a question yet to be answered.

While developing the right crowdsourcing task is necessary for identifying SMOs in online social movements, it is not sufficient. It would be costly and time consuming to obtain human judgment on millions of Twitter accounts. Therefore, we developed reliable classifiers with high accuracy to perform this task at scale. Finally, the research team has developed a clustering method to classify SMOs into social movement industries (SMIs) in an unsupervised fashion. This method identified organizations from 12 different SMIs engaged in the BlackLivesMatter movement. The labels collected and classifiers built, and the clustering technique implemented provide a great opportunity for future research in organization theory and collective action to study SMOs at a scale never possible before. As such, we will share the data publicly to help advance the field.

[25] Within the scope of our dataset, *blklivesmatter*, the official Twitter account for the movement, has 34K followers, whereas the current number is at 249K.

[26] Here, the *x axis* represents the number of days since the start of the movement, and the *y axis* denotes the number of tweet contributions each day. The colored lines are the fitted regression lines $y \sim x$ of each SMI.

REFERENCES

[1] Alexa. 2017. Find Website Traffic, Statistics, and Analytics. (2017). http://www.alexa.com/siteinfo [Online; accessed 24-April-2017].

[2] Giselle A Auger. 2013. Fostering democracy through social media: Evaluating diametrically opposed nonprofit advocacy organizationsâĂŹ use of Facebook, Twitter, and YouTube. *Public Relations Review* 39, 4 (2013), 369–376.

[3] Mathieu Bastian, Sebastien Heymann, Mathieu Jacomy, et al. 2009. Gephi: an open source software for exploring and manipulating networks. *Icwsm* 8 (2009), 361–362.

[4] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* 2008, 10 (2008), P10008.

[5] Phillip Bonacich. 1987. Power and centrality: A family of measures. *American journal of sociology* 92, 5 (1987), 1170–1182.

[6] Andrew P Bradley. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition* 30, 7 (1997), 1145–1159.

[7] L Breiman. 2001. Random Forests Machine Learning. 45: 5–32. *View Article PubMed/NCBI Google Scholar* (2001).

[8] Jon Bruner. 2013. Tweets Loud and Quiet. (2013). https://www.oreilly.com/ideas/tweets-loud-and-quiet

[9] Axel Bruns and Jean E Burgess. 2011. The use of Twitter hashtags in the formation of ad hoc publics. In *Proceedings of the 6th European Consortium for Political Research (ECPR) General Conference 2011*.

[10] Ceren Budak, Sharad Goel, and Justin M. Rao. 2016. Fair and Balanced? Quantifying Media Bias through Crowdsourced Content Analysis. *Public Opinion Quarterly* 80, S1 (2016), 250.

[11] Michael Buhrmester, Tracy Kwang, and Samuel D Gosling. 2011. Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? *Perspectives on psychological science* 6, 1 (2011), 3–5.

[12] Chris Callison-Burch. 2009. Fast, cheap, and creative: evaluating translation quality using Amazon's Mechanical Turk. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*.

Association for Computational Linguistics, 286–295.
[13] Zi Chu, Indra Widjaja, and Haining Wang. 2012. Detecting social spam campaigns on twitter. In *International Conference on Applied Cryptography and Network Security*. Springer, 455–472.
[14] Thomas Cover and Peter Hart. 1967. Nearest neighbor pattern classification. *IEEE transactions on information theory* 13, 1 (1967), 21–27.
[15] Lindley Curtis, Carrie Edwards, Kristen L Fraser, Sheryl Gudelsky, Jenny Holmquist, Kristin Thornton, and Kaye D Sweetser. 2010. Adoption of social media for public relations by nonprofit organizations. *Public Relations Review* 36, 1 (2010), 90–92.
[16] Dmoz. 2002. Open directory project. (2002).
[17] Nadbor Drozd. 2016. Text Classification With Word2Vec. http://nadbordrozd.github.io/blog/2016/05/20/text-classification-with-word2vec/. (2016).
[18] Jennifer Earl. 2015. The future of social movement organizations: The waning dominance of SMOs online. *American Behavioral Scientist* 59, 1 (2015), 35–52. http://journals.sagepub.com/doi/abs/10.1177/0002764214540507
[19] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. LIBLINEAR: A library for large linear classification. *Journal of machine learning research* 9, Aug (2008), 1871–1874.
[20] Roberto M. Fernandez and Doug McAdam. 1988. Social Networks and Social Movements: Multiorganizational Fields and Recruitment to Mississippi Freedom Summer. *Sociological Forum* 3, 3 (1988), 357–382. http://www.jstor.org/stable/684338
[21] Tina Fetner and Brayden G King. 2016. Three-Layer Movements, Resources, and the Tea Party. *Understanding the Tea Party Movement* (2016), 35.
[22] Ryan J Gallagher, Andrew J Reagan, Christopher M Danforth, and Peter Sheridan Dodds. 2016. Divergent discourse between protests and counter-protests:# BlackLivesMatter and# AllLivesMatter. *arXiv preprint arXiv:1606.06820* (2016).
[23] Pierre Geurts, Damien Ernst, and Louis Wehenkel. 2006. Extremely randomized trees. *Machine learning* 63, 1 (2006), 3–42.
[24] Sandra González-Bailón. 2013. Online social networks and bottom-up politics. (2013). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2246663
[25] Sandra González-Bailón, Javier Borge-Holthoefer, Alejandro Rivero, and Yamir Moreno. 2011. The dynamics of protest recruitment through an online network. *Scientific reports* 1 (2011).
[26] Chao Guo and Gregory D Saxton. 2014. Tweeting social change: How social media are changing nonprofit advocacy. *Nonprofit and Voluntary Sector Quarterly* 43, 1 (2014), 57–79.
[27] Asmelash Teka Hadgu, Kiran Garimella, and Ingmar Weber. 2013. Political hashtag hijacking in the US. In *Proceedings of the 22nd International Conference on World Wide Web*. ACM, 55–56.
[28] Matthew Honnibal and Mark Johnson. 2015. An Improved Non-monotonic Transition System for Dependency Parsing. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Lisbon, Portugal, 1373–1378. https://aclweb.org/anthology/D/D15/D15-1162
[29] Sarah J Jackson and Brooke Foucault Welles. 2015. Hijacking# myNYPD: Social media dissent and networked counterpublics. *Journal of Communication* 65, 6 (2015), 932–952.
[30] Zaid Jilani. 2011. REPORT: Five Things Unions Have Done For All Americans. (2011). https://thinkprogress.org/report-five-things-unions-have-done-for-all-americans
[31] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2016. Bag of Tricks for Efficient Text Classification. *arXiv preprint arXiv:1607.01759* (2016).
[32] Darius Kazemi. 2015. Occupations. https://github.com/dariusk/corpora/blob/master/data/humans/occupations.json. (2015).
[33] Ernest Sydney Keeping. 1962. *Introduction to statistical inference*. Courier Corporation.
[34] J Peter Kincaid, Robert P Fishburne Jr, Richard L Rogers, and Brad S Chissom. 1975. *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel*. Technical Report. Naval Technical Training Command Millington TN Research Branch.
[35] Terry Koo and Michael Collins. 2010. Efficient third-order dependency parsers. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 1–11.
[36] Kristen Lovejoy, Richard D Waters, and Gregory D Saxton. 2012. Engaging stakeholders through Twitter: How nonprofit organizations are getting more out of 140 characters or less. *Public Relations Review* 38, 2 (2012), 313–318.
[37] Fiona Martin and Mark Johnson. 2015. More efficient topic modelling through a noun only approach. In *Australasian Language Technology Association Workshop 2015*. 111.
[38] Winter Mason and Siddharth Suri. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods* 44, 1 (2012), 1–23.
[39] Doug McAdam. 1986. Recruitment to High-Risk Activism: The Case of Freedom Summer. *Amer. J. Sociology* 92, 1 (1986), 64–90. http://www.jstor.org/stable/2779717
[40] Doug McAdam, John D McCarthy, Susan Olzak, and Sarah A Soule. 2009. Dynamics of collective action. https://web.stanford.edu/group/collectiveaction/cgi-bin/drupal/. (2009).
[41] John D. McCarthy and Mayer N. Zald. 1977. Resource mobilization and social movements: A partial theory. *American journal of sociology* 82, 6 (1977), 1212–1241. http://www.journals.uchicago.edu/doi/abs/10.1086/226464
[42] John D. McCarthy and Mayer N. Zald. 1977. The trend of social movements in America: Professionalization and resource mobilization. (1977). https://deepblue.lib.umich.edu/handle/2027.42/50939
[43] Seungahn Nah and Gregory D Saxton. 2013. Modeling the adoption and use of social media by nonprofit organizations. *New Media & Society* 15, 2 (2013), 294–313.
[44] Janine Nahapiet and Sumantra Ghoshal. 1998. Social Capital, Intellectual Capital, and the Organizational Advantage. *The Academy of Management Review* 23, 2 (1998), 242–266. https://doi.org/10.2307/259373
[45] A Conrad Nied, Leo Stewart, Emma Spiro, and Kate Starbird. 2017. Alternative Narratives of Crisis Events: Communities and Social Botnets Engaged on Social Media. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 263–266.
[46] Jonathan A Obar, Paul Zube, and Clifford Lampe. 2012. Advocacy 2.0: An analysis of how advocacy groups in the United States perceive and use social media as tools for facilitating civic engagement and collective action. *Journal of information policy* 2 (2012), 1–25.
[47] Alexandra Olteanu, Ingmar Weber, and Daniel Gatica-Perez. 2015. Characterizing the demographics behind the# blacklivesmatter movement. *arXiv preprint arXiv:1512.05671* (2015).
[48] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
[49] James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates* 71, 2001 (2001), 2001.
[50] John Platt et al. 1999. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers* 10, 3 (1999), 61–74.
[51] Philipp Schaer. 2012. Better than their reputation? on the reliability of relevance assessments with students. In *International Conference of the Cross-Language Evaluation Forum for European Languages*. Springer, 124–135.
[52] Sarah A Soule. 2012. Social movements and markets, industries, and firms. *Organization Studies* 33, 12 (2012), 1715–1733.
[53] Emma S Spiro and Andrés Monroy-Hernández. 2016. Shifting Stakes: Understanding the Dynamic Roles of Individuals and Organizations in Social Media Protests. *PLOS ONE* 11, 10 (Oct. 2016), e0165387. https://doi.org/10.1371/journal.pone.0165387
[54] Suzanne Staggenborg and Josee Lecomte. 2009. Social movement campaigns: Mobilization and outcomes in the Montreal women's movement community. *Mobilization: An International Quarterly* 14, 2 (2009), 163–180. http://mobilizationjournal.org/doi/abs/10.17813/maiq.14.2.04l4240734477801
[55] J Tang, S Alelyani, and H Liu. 2014. Data classification: algorithms and applications. *Data Mining and Knowledge Discovery Series, CRC Press* (2014), 37–64.
[56] Verta Taylor. 1989. Social Movement Continuity: The Women's Movement in Abeyance. *American Sociological Review* 54, 5 (1989), 761–775. https://doi.org/10.2307/2117752
[57] Robert L Thorndike. 1953. Who belongs in the family? *Psychometrika* 18, 4 (1953), 267–276.
[58] David H Wolpert. 1992. Stacked generalization. *Neural networks* 5, 2 (1992), 241–259.
[59] Mayer N. Zald and Roberta Ash. 1966. Social Movement Organizations: Growth, Decay and Change. *Social Forces* 44, 3 (1966), 327–341. https://doi.org/10.2307/2575833
[60] Mayer N. Zald and John D. McCarthy. 1979. Social movement industries: Competition and cooperation among movement organizations. (1979). https://deepblue.lib.umich.edu/handle/2027.42/50975
[61] Jinhong Zhong, Ke Tang, and Zhi-Hua Zhou. 2015. Active Learning from Crowds with Unsure Option.. In *IJCAI*. 1061–1068.
[62] Stephen Zunes and Jesse Laird. 2010. The US Civil Rights Movement (1942-1968). https://www.nonviolent-conflict.org/the-us-civil-rights-movement-1942-1968/. (2010).