Beyond the Eye-Catchers: a Large-Scale Study of Social Movement Organizations' Involvement in Online Protests

Proponents of Internet activism postulate that social media affordances drastically reduce education, coordination and participation costs, and thus render social movement organizations (SMOs) nonessential. Yet, such claims currently lack sufficient evidence due to the difficulty of identifying SMOs at scale. This paper aims to fill that gap and characterize the role SMOs play in online protest movements.

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 movements. Our results reveal that beyond the few high profile eye-catcher SMOs, the rest vary significantly in terms of role and importance, and that an average SMO behaves surprisingly similar to an average individual. We set out to unearth the reasons behind this finding. By assigning each SMO to its respective social movement industry (SMI) according to its primary objectives using unsupervised methods, we show that an extensive number of SMOs from peripheral SMIs participated, alongside SMOs from the core SMI, most commonly in solidarity with lower commitment. Finally, we propose a predictive model of SMOs mobilization volume. Our analysis shows that structural embeddedness of SMOs within the SMO network is the biggest indicator of success—SMOs that have more bi-directional ties (or, alliances) with other SMOs have higher mobilization volume. Lastly, SMOs that have higher network brokerage or access to more varied information also possess higher mobilization volume.

ACM Reference Format:

. 2018. Beyond the Eye-Catchers: a Large-Scale Study of Social Movement Organizations' Involvement in Online Protests. 1, 1 (April 2018), 18 pages. https://doi.org/0000001.0000001

1 INTRODUCTION

The importance of highly professionalized and respected social movement organizations (SMOs) have been extensively documented in prior studies. Indeed, SMOs advanced the recruitment of participants in high risk activism [25]. SMOs provide the structure and leadership needed to coordinate different groups of participants for swift offline actions [63]. Finally, SMOs sustained movements and kept their members' ideologies alive in times of public apathy or hostility [68].

Yet, Internet activism slowly replacing traditional civic engagements has brought into question the significance of SMOs in the era of Web 2.0. Some scholars argue that the decentralized structure of the Web and social media affordances have drastically lowered the associated costs, rendering SMOs less important [7, 8, 22, 23, 40, 61]. Earl and Kimport show that many online protests such as petitions and boycotts are organized by individuals [23]. Bennett and Segerberg, studying Arab Spring and Occupy Wall Street, make similar observations [7]. This conclusion is, however, not uniformly accepted. For instance, Fetner and Key [27], by analyzing the role of teapartypatriots.org in the Tea Party movement, credit the movement's success to the well financed interest groups behind it. Spiro et al. [62] show that SMOs eventually supersede any one individual activist as movements progress and become the center for broadcasting information and organizing events. In other words, they play a crucial role without which the success of a movement would be short-lived [68].

Author's address:

XXXX-XXXX/2018/4-ART $15.00

https://doi.org/0000001.0000001

There is a growing academic interest in Internet activism. [5, 15, 20, 30, 64, 69, 72]. Given SMOs' importance in past social movements and the incongruent hypotheses on their role in the new era, it's surprising that only few distinguish SMOs' participation from individuals and other organizations. The ones that do make the distinction [34, 45, 56] often times require substantial manual labor and only focus on the eye-catchers—a limit set of exceptionally successful and popular SMOs.

Unlike the aforementioned studies, we move beyond focusing on selected few high profile SMOs and provide the first automated method in identifying SMOs at scale. Using this approach, we identify and study the behavior of more than 50 thousand SMOs in 2 distinct *online social media protests*—the BlackLivesMatter movement and gender-related advocacies broadly categorized as Women's rights movement. We provide the following contributions:

(1) Consistent with prior studies [27, 62], eye-catcher SMOs are indeed more committed to protests; they partake in more knowledge sharing and community building efforts, and draw significant fraction of endorsements and acknowledgments. For instance, the top 5% quantile SMOs received ≥ 135 retweets, while the number is 5 and 15 for the top 5% quantile individuals and other organizations respectively.
(2) More interestingly, however, we observe that an average SMO is very similar to an average individual both in terms of commitment and importance. For instance, over half of all SMOs in BlackLivesMatter only contributed 1 or 2 protest tweets and had 0 recruitment. The similarity is even stronger for individuals that are structurally close to SMOs. This demonstrates the existence of a long tail of SMOs beyond the eye-catchers studies traditionally concentrated on.
(3) Focusing on the observed variance, we first show that it is partially due to differences in social movement industries (SMI). Indeed, an SMO with the goal of defending women's reproductive rights can easily show solidarity to an online movement of another cause, while spending the majority of its efforts on Women's Rights movement. We identify SMIs using automated unsupervised methods and show that SMOs from an extensive set of peripheral SMIs, including *LGBTQ*, *Women&Health* and 9 others, protested in solidarity in BlackLivesMatter alongside the SMOs from the core *African American* SMI. Further, we observe that some of the "peripheral" SMIs with established networks and resources played a vital role in the development of the BlackLivesMatter movement despite having other primary objectives. Additionally, we also see a significantly lower level of solidarity for Women's Rights: only 2 other SMIs protested alongside the core *Women* SMI.
(4) Finally, we build predictive models of SMOs mobilization volume operationalized using the following metrics: i) the number of times when an activist's first protest tweet is a retweet of or mentions an SMO, and ii) the number of an SMO's followers that participated in a movement *after* the first protest tweet contributed by the SMO. We identify variables including account characteristics, network attributes, strategies, and participation patterns that are indicative of how well an SMO can successfully induce its followers and others to protest. Our models have moderate to substantial predicitive power for both metrics: adjusted R^2 of 0.52 and 0.92 respectively with cross validation.

In sum, this study presents the first large-scale analysis of SMOs for online movements; it exposes a long tail of SMOs, substantial in number, beyond the eye-catchers. We illustrate characteristics and behaviors of these SMOs and highlight the various important implications of this tail.

2 BACKGROUND & RELATED WORK

We draw from numerous classic theories and studies addressing SMOs within traditional participation media [25, 48, 51, 54, 68, 73, 74]. We adopt the the definition of social movement organization from the Resource Mobilization Theory proposed by McCarthy and Zald [51, 73]. In addition, we observe and adhere to the coding schemes published by McAdam [49]. We define the following:

Social Movement Organizations (SMO): 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 other organizations in two ways. First, SMOs do not provide regular goods or services[1]. Second, SMOs' incentives are primarily purposive. 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 [73]. Examples of SMOs are shown in table 1.

Social Movement Industries (SMI): An SMI is a group of SMOs that share broad preferences or goals of a social movement [50]. Examples of SMIs include *Christianity* and *Environment*[3].

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

Individuals: Twitter accounts that are neither SMOs nor OTHOs will be coded as individuals.

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

There are several recent papers on SMOs [17, 22, 27, 34, 45, 53, 56, 62] that are worth noting here. Spiro et al. [62] examine the distinctions between zealots and SMOs in initializing and sustaining online social movements. Conversely, research conducted by Lovejoy et al. [45] suggests that SMOs fell short of employing social media optimally. Furthermore, analyses done by Earl [23], Karpf [40], and Shirky [61] suggest the waning importance of SMOs due to social media affordances and personalized political participation. Unlike the aforementioned studies, our research is based on a more comprehensive set of SMOs.

3 DATA ACQUISITION & ANNOTATION

Our dataset consists of over 80 hashtag protests that are categorized into 2 broad topics: the BlackLivesMatter movement and Women's Rights movement. There are over 50 raced-related hashtag movements including #ferguson, #blacklivesmatter, #policebrutality, et cetera, spanning 36.6 million

[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.

[3]For instance, *Greenpeace* and *Sierra Club* belong to this SMI.

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, women's healthcare, women empowerment, wage equality, et cetera. The dataset includes *all* tweets with the corresponding hashtags between 02-2014 and 05-2015.

Given Twitter does not supply account categories, we use crowdsourcing and supervised learning to classify all users into 3 distinct groups: SMOs, OTHOs, and individuals.

3.1 Crowdsourcing Task

We recruit MTurk workers to label a subset of Twitter accounts into SMOs, OTHOs, and individuals.
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 OTHOs 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 or DMOZ[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 OTHOs) while still sampling randomly from the broader distribution for testing purposes. In addition, we also obtain the dataset from Olteanu et al. [57] which contain over 300 labeled organization accounts. Our final selected set has approximately 2000 accounts.
Instructions: We take screenshots of these accounts' Twitter page, 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, we first provide workers the definitions of SMOs, examples, and instructions for inclusions, exclusions, and exceptions. Next, we take them through a training task before finally starting on the actual task [7].
Qualifications and Monitoring: Success of a crowdsourcing task relies heavily on the right mechanisms to ensure worker qualifications [47]. We require that workers: 1) reside in the U.S. 2) have successfully completed at least 1,000 HITs; and 3) have an approval rate of at least 98%. We first provide workers with instructions on how to identify SMOs, we then take them through a training process including 10 pre-labeled accounts. Furthermore, we allow workers to skip cases they are unsure about while still providing compensation[8]. Such choices are shown to improve label quality [75]. Finally, we inject a number pre-labeled Twitter accounts as a gold standard, monitor worker accuracy for such cases and block workers whose accuracy fall below 60%.
Evaluation: Despite the complexity of the task, the workers provide high quality and reliable labels with sufficient training and quality checks. We apply Krippendorff's alpha separately to assess inter-rater reliability. Our labels have a scores of 0.72 indicating a substantial agreement [60]. In addition, we also calculate Turkers' Gold set (pre-labeled by the researchers) accuracy at 82%.

[4]Threshold is informed by related work on Twitter accounts [13] as the 99 percentile in terms of followers.
[5]We crawl the homepages of all the websites listed under *Society* and *News* for both Alexa and DMOZ, and gather all the Twitter users contained within these homepages. Our crawler obtain 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 3.7% of tasks did the workers choose this option.

3.2 Supervised Classification Task

There are several existing methods that focus on categorizing Twitter accounts [19, 42, 52, 58], many of which are binary classifications (organizations v.s. individuals). Unlike these related work, we seek to further classify organizations into SMOs and OTHOs. For that, 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 an OTHO. Such a nested classification approach has been shown to work well in the case of imbalanced classes [14]. This approach also allows us to identify the optimal combinations of featuresets, preprocessing steps and classifier types, which yield better results than a regular multiclass classification method.

Table 2. Feature Selection

Feature-set	Features
User Name& User Handle	(i) Whether field contains an actual person's name (python NLTK Name Corpus); (ii) whether user handle is included within any hyperlinks in the homepages of websites listed under Alexa and Dmoz's *News* and *Society* categories.
User Profile	Whether field contains (i) an URL, (ii) an occupation (Kazemi's Occupation Corpus[41]), (iii) numbers, (iv) nonascii characters (e.g. emojis); (v) whether URLs, if present, belong to a social media website; (vi) readability (Flesch–Kincaid Grade), (vii) word semantic vectors in the dimensions of 10, 25, and 50 using Fasttext's Skipgram method [21, 39], (viii) word dimensions, such as *pronoun*, using LIWC [59].
User Motive	(i) Total tweet count, (ii) protest tweet count, (iii) protest retweet count, (iv) protest mention count, (v) ratio between (i) and (ii), (vi) number of days between users' first and last protest tweets, (vii) number of distinct protest hashtags, (viii) the timestamp of users' first protest tweet
User Influence	(i) Friend count, (ii) follower count, (iii) ratio between (i) and (ii), (iv) number of retweets and mentions users received from others. All in log-scale.
Tweets	(i) Number of original protest tweets (i.e. not a retweet), (ii) url count, (iii) hashtag count, (iv) non-ascii characters count, (vi) readability. (vii) word semantic vectors using Fasttext's Skipgram, (viii) word dimensions using LIWC.

Featureset Selection: In Table 2, we list our feature selections with the intent to discriminate individuals from organizations and/or SMOs from OTHOs.
Classifier Selection: We select 5 commonly used *single classifiers*: Linear Regression, Linear Support Vector Machine (SVM),[9], Random Forest, Extra Trees, and K-Nearest Neighbor as implemented by Python's *Scikit-learn* library to compute baseline performance. We generate $5C3$ (5 choose 3) combinations of classifiers from the single classifier list and assign the selected 3 as base estimators to a *voting ensembler* as implemented by *Scikit-learn*. We apply another ensemble classification technique called *Stacking* [66] 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.
Classification Performance: 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 assess the best performing classification pipeline using the sum of the minority class's precision and recall.
Individual and Organization Classification: Detailed evaluation statistics are summarized in Table 3[10]. All our classifiers for both test-set types have AUC above 0.9 which is considered *excellent* for an unbalanced datasets [11]. 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 with 0.93

[9]We apply Platt's scaling to derive the prediction probability.
[10]Due to space limitations, we only present results for the best preforming classifier from each category for the minority class.

precision and 0.68 recall for classifying organizations [11] Using this classifier, we categorize 312K accounts in our dataset as organizations. We also inspect the features with the highest weights to gain insights about the classifier. Unsurprisingly, having an actual person's name or an occupation in the profiles strongly indicate accounts as individuals.

SMO and OTHO 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. 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 more by other participants suggest accounts as SMOs instead of OTHOs. We apply the best performing classifier to the set of 312K accounts classified as organization and identify more 50 thousand accounts as SMOs [12].

Table 3. Classification Summary

		Individual/Organization				SMOs/OTHOs			
		accuracy	precision	recall	auc	accuracy	precision	recall	auc
strata	type		org	org			smo	smo	
all	stack	0.97	0.98	0.93	0.98	0.87	0.83	0.76	0.83
all	voting	0.92	0.94	0.83	0.95	0.86	0.86	0.71	0.82
all	single	0.91	0.93	0.84	0.94	0.84	0.85	0.66	0.84
random	stack	0.98	0.93	0.69	0.94	0.88	0.68	0.69	0.77
random	voting	0.98	1.0	0.5	0.97	0.89	0.79	0.56	0.77
random	single	0.98	0.92	0.68	0.96	0.88	0.75	0.54	0.77

4 ANALYSIS

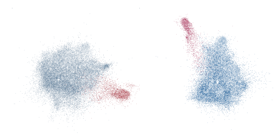
We first conduct additional quality assessment by addressing Twitter bots, Countermovement SMOs, and robustness check. We then explore the importance of SMOs in comparison to individuals and OTHOs in Section 4.2. We see a substantial variance in SMOs role in online protests. Next, we provide a partial explanation of the variance observed by examining SMOs within the context of SMIs in Section 4.3. Finally, we operationalize SMOs mobilization volume using 2 metrics and propose models robust and predictive of their mobilization performance in Section 4.4.

4.1 Quality Assessment

Automated v.s Genuine Accounts: There is a growing body of research addressing how interest groups use bot farms to both support and interfere with social movements [26, 29, 35]. To measure the "genuineness" of the SMO accounts, we first select the subset of SMO accounts that belong to the top 5 percentiles in terms of protest tweet contribution(18 or more tweets). We then use the *Botometer* API [71] to assign each account a bot score. However, scores for organization accounts—often operated by multiple individuals, or combinations of users and automatic tools—are commonly

[11] We qualitatively examine organization accounts that were incorrectly classified as individuals; we observe that 42.8% left their user profile description field blank.

[12] To ensure a more accurate assessment of SMOs' role in social media activism, we also examine the most active (>=99.99 percentile in number of protest tweets contributed), or most influential(>=99.99 percentile in both number of retweets and mentions received) participants within our dataset. We found over 700 highest value participants within our labeled organizations dataset and manually reassign them into correct categories as SMOs, OTHOs, or individuals. Accuracy measures for this sample is comparable to the findings in the test set. For instance, approximately 78% of SMOs are correctly identified by our classifier.



(a) BLM (b) Women

Fig. 1. Countermovement SMOs.

inaccurate[13]. This shortcoming is indeed acknowledged by [71]. Consequently, we choose to include all SMOs for analyses regardless of bot score. Nevertheless, as a robustness check, when comparing SMOs' direct tweet contribution to OTHOs and individuals, we repeat the analysis excluding the subset of SMOs with a bot score of > 0.50; observations are consistent.

Countermovement SMOs: Previous research shows that countermovement SMOs utilize strategies such as alternative narrative or framing [55], hashtag hijacking and spamming [38], and spreading counter-protest hashtags [32] to attack and discredit the "offending" movement. To ensure the SMOs under examination are in fact supportive of BlackLivesMatter and Women's Rights, we 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 exists if u is followed by v.[14] We apply the same procedure to generate G_{women}. Graphs are given in Figure 1[15]. Both graphs 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 randomly sample 100 tweets from BlackLivesMatter, each posted by a distinct account from the red subset, and manually label them as supportive or opposing. We note that 79 tweets objected the movement. We conduct a similar analysis on Women's Rights and observe a comparable pattern. We remove these accounts for the subsequent analyses.

Robustness Check: To ensure our observations are not simply due to individuals' accounts being misclassified as SMOs, as a robustness check, we use Youten's Index [28] to select optimal probability cut-off values for our classifiers to limit the false positive rate to only 2% for accounts belonging to the random strata.[16] Results obtained using the 26K and 50K SMOs datasets are consistent.

4.2 Comparing SMOs' to Other Organizations and Individuals

We observe that 33.7K and 19.6K SMOs participated in BlackLivesMatter and Women's Rights respectively. Despite being a very small fraction of all accounts, SMOs contributed a substantial fraction of tweets, and received a considerable fraction of endorsements (retweets), and acknowlegements (mentions). In this section, we go beyond direct contributions by providing a more in-depth comparison of SMOs' involvement in online protests to that of OTHOs and individuals using the following 5 metrics: *commitment*, *knowledge sharing*, *community building*, *structural significance*, and *recruitment*. We select these metrics by adopting from existing literature focused on SMOs online strategies [3, 34, 56] and related network cascading theory [33]. Here, we also separate individuals

[13] Other well known SMOs with a bot score above 0.5 include *unicefusa (United Nations Children's Fund in U.S.), greenpeaceusa, naacp, occupywallst, uniteblue, mmfa, seiu, hrw, ufcw, et cetera*.

[14] We use Twitter's streaming API to collect SMOs friends and followers.

[15] We use Gephi's[6] build-in graph layout and network clustering features: *ForceAltas2*, and *Modularity*

[16] Here, an account from random strata is labeled as an SMO only if the classifier has a prediction probability of >= .79

Table 4. SMOs Selected Contribution Overview

Dimension	Measurements	BLM SMO	BLM OTHO	BLM Indv(H)	BLM Indv	Women SMO	Women OTHO	Women Indv(H)	Women Indv
Commitment	mean(median) fraction of accounts active per day	2.28% (1.60%)	1.19% (0.52%)	1.46% (0.58%)	0.52% (0.12%)	0.97% (0.68%)	0.49% (0.25%)	0.59% (0.27%)	0.33% (0.12%)
Structural Significance	mean(median) kcore	32.13(4)	14.13(3)	16.04(4)	4.33(2)	7.45(3)	4.1(2)	4.63(2)	2.59(2)
Knowledge Sharing	mean(median) number of URLs	40.94(3)	8.74(2)	10.44(2)	3.05(1)	6.46(2)	2.78(1)	2.94(1)	1.58(1)
Community Building	mean(median) number of protest communities	1.83(1)	1.65(1)	1.81(1)	1.34(1)	1.26(1)	1.13(1)	1.19(1)	1.05(1)
Recruitment	total number (fraction) of recruits	355.84K (10.76%)	585.67K (17.68%)	983.34K (29.68%)	1.4M (41.9%)	458.88K (18.35%)	166.91K (6.67%)	390.29K (15.60%)	1.48M (59.37%)
	mean(median) number of recruits	10.55(0)	3.62(0)	1.63(0)	0.37(0)	23.39(0)	2.06(0)	1.2(0)	0.72(0)

who are structurally close to SMOs (having at least 1 bi-directional tie with an SMO) from the rest. We denote these individuals as indv(H), with H representing high structural closeness to SMOs, and they serve as an additional baseline comparison group.

Commitment For each group, we compute the fraction of accounts actively participated per day averaged over time. We also compute the number of days elapsed between the first and last tweet posted by an account (i.e. protest length) and the number of unique days an account protested.

Knowledge Sharing Previous studies on SMOs' social media strategies identity the use of URLs as a knowledge sharing strategy[3, 34, 56]. We compute the total number of unique URLs and domains shared by each group; we also calculate the number of URLs shared by each account.

Community Building Existing literature suggests SMOs use hashtags more frequently than the baseline usage on Twitter for community-building purposes [34]: they can be discovered more easily through hashtag based searches on Twitter and bring interested parties together. To evaluate whether SMOs indeed reach out to more protest communities, we construct hashtag-hashtag co-occurrence graphs and cluster similar hashtags into communities. [17]. We then derive the number of communities each group chose to focus on.

Structural Significance We use two metrics to assess SMOs' structural significance: 1) *k-core* score, which measures a participant's ability in recruitment and protest cascading [33], and 2) *in-degree*, which captures the participant's reputation [10]. 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, and an edge $e_{u,v} \in E_{blm}$ if u mentions or retweets v at least once (similar construction for G_{women}).[18] We compute the in-degree and k-core value of nodes corresponding to SMOs, OTHOs, and individuals.

Recruitment 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 and if a' had already been participating in the movement(i.e.

[17]We define 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 [9] and identify 7 communities for BlackLivesMatter and 5 communities for Women's Rights. Sample Hashtags belonging to the same community are #wearesilent and #wearehere.

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

timestamp of $t_{a'_{blm}}$ is less than $t_{a_{blm}}$), we denote a' as the recruiter of a into the corresponding movement[19]. Then we aggregate the total number of recruitments for each account a'_{blm}.

Results. Measurements from each dimension are summarized in Table 4. Due to space limitation, only a representative subset of the metrics are selected; the omitted measures are consistent with what's shown. We observe that, on average, SMOs indeed play a more significant role in online protests. For example, in BlackLivesMatter, SMOs on average have a kcore value of 32, which is 2 times that of indv(H) despite their structural closeness, and 5 times that of what is observed for individuals on average. The average number of recruits by BlackLivesMatter SMOs is 11, and OTHOs is 4 and only 2 for indv(H).

More interestingly, however, we see that the averages summarized below are particularly high for SMOs due to the small fraction of highly influential accounts. For instance, the top 5% quantile SMOs received ≥ 135 retweets, while the number is 5 and 15 for the top 5% quantile individuals and OTHOs respectively. Indeed, an average SMO (characterized by medians listed in Table 4) looks surprisingly similar to an average individual. For instance, an average BlackLivesMatter SMO only contributed 1 or 2 protest tweets, is structurally insignificant to the movement, and has 0 recruitment. Our study reveals the existence of a long tail of SMOs that studies only focus on eye-catchers may have missed. Our subsequent analysis focuses on unpacking the characteristics of these long tail SMOs and their mostly "supportive" role in online protests.

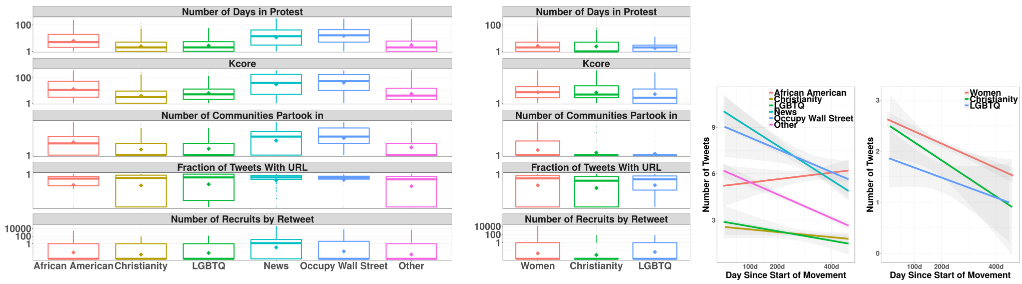
Table 5. Samples of SMI Clustering, Precision, and Profile

SMI Cluster	Precision	Sample Account Profile
Christianity (BLM)	0.96	1.these are updates, information and prayers for the #downtown congregation of #redeemer 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. 3.hello from westside unitarian universalist church! we're proud to be a progressive voice in ft worth. check out our website & drop in for a visit.
Occupy Wall Street (BLM)	0.68	4.in solidarity with occupy movements worldwide. #r##n# non violent. pro justice 5.occupy, politics, anti-corruption, anti-megacorp, against dangerous levels of power in any form. for the people. 6.west hollywood division of occupy los angeles. be on the right side of history. stands in solidarity with #ows #ola.
African American (BLM)	0.92	7.black is beautiful#n#black lives matter#n#civil rights for all#n#love unconditionally 8.#blacklivesmatter is an affirmation and embrace of the resistance and resilience of black people. 9.an auxiliary organization on college campuses nationwide directly related to, and in turn shares the ideas and views of, the 100 black men of america.
Women (Women)	0.80	10.world pulse is the leading network using the power of digital media to connect women worldwide and bring them a global voice. 11.exposing corrupt cops in frisco texas. we believe the constitution is more than a piece of paper. we expose social media frauds, child killers/molesters. 12.The country's largest provider of specialist services for women and children escaping domestic violence. Twitter account staffed Mon-Fri 9.30-5.30.
LGBTQ (Women)	0.96	13.mygwork is an online lgbt professional network. we connect individuals, gay and lesbian friendly companies, associations and local businesses. 14.michigan's only statewide anti-violence and advocacy organization serving the lesbian, gay, bisexual, and transgender communities for over 20 years. 15.re-defining a civil rights movement for the 21st century. gay lesbian bisexual transgender queer lgbt glbt lgbtq

4.3 Comparing SMOs from Different SMIs

One likely explanation for the considerable variance in SMOs involvement in online protests is that primary objectives, thus SMIs, of SMOs vary widely. Here, we provide detailed analyses of the role differences between SMOs from the core SMI and the ones from peripheral SMIs.

[19]If the first tweet by a_{blm} contains multiple mentions and/or retweets, each account is considered a recruiter.



(a) BLM (b) Women (c) BLM (d) Women

Fig. 2. Figure 2(a) and 2(b) depict differences in role fulfillment between SMOs from distinct SMIs including *commitment, structural significance, knowledge sharing, community building* and *recruitment*. Figure 2(c) and 2(d) highlight their change in protest intensity as movements continue.

4.3.1 Categorize SMOs into SMIs. We first apply part-of-speech (POS) tagging to each SMO's profile description.[20] We focus on noun tokens given their strength in extracting semantically meaningful topics [46]. 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. In addition, we manually examine the most frequently occurring non-noun tokens and selectively keep the ones that are contextually relevant.[21] This results in dictionaries of size 1.7K and 1.6K for BlackLivesMatter and Women's Rights respectively. We keep the set of accounts that have 4 or more relevant tokens in their profiles and discard the rest, resulting in 22K matching accounts for BlackLivesMatter and 15K for Women's Rights. For each movement, we first translate tokens of the remaining accounts into a sparse document term matrix and apply the kmeans elbow method [70] to determine the optimal number of document clusters; we then use kmeans clustering[22]. Finally, we randomly select a sample of 25 accounts from each of the resulting clusters and manually assess their primary objectives.

Using this approach, we identify 12 distinct SMIs for BlackLivesMatter, and 3 for Women's Rights[23]. Randomly selected accounts from a representative subset of SMIs and the corresponding sample precisions are listed in Table 5. We will first examine SMIs in BlackLivesMatter follow by Women's Rights. For both, we reapply our analyses from the previous section including *commitment, knowledge sharing, community building, structural significance* and *recruitment* to assess, with SMI context, SMOs' role as the educator, recruiter, coordinator, and sustainer for the movements.

4.3.2 BlackLivesMatter. Alongside the SMOs of the core *African American* SMI, an extensive set of SMOs from *Youth, Student, Christianity, Research Institutes, LGBTQ, Social Welfare, Non-African American Minorities, Occupy Wall Street (OWS)* and *News* [24] SMIs participated to support BlackLivesMatter, highlighting the strength of social media in enabling collaboration between SMOs with vastly different primary objectives. Here, we select a representative subset of the SMIs: *African American, Christianity, LGBTQ, News, OWS*, and aggregate all others into the *Other* category.

[20]We apply *spaCy* [36].

[21]Sample non-noun tokens we keep include"black"—an adjective, but highly relevant to the context; "occupy", normally a verb, but here signals relation to the Occupy movement.

[22]We explore the use of LDA (latent dirichlet allocation) and have find the topics to be less coherent. We also try 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 is also unfruitful.

[23]All identified SMIs have a random sample precision of approximately 70% or above.

[24]Note that *News* SMOs differ from mainstream corporate media firms in that *News* SMOs center their objectives around shining light into social, political, or economic issues, and/or bringing about positive depictions of certain groups.

As shown in Figure 2, SMO accounts belonging to the *African American* SMI, denoted as A_{aa}, are more central to BlackLivesMatter than the subset of SMOs that participated in solidarity such as $A_{christianity}$ and A_{lgbtq}. The median tweet contribution by A_{aa} is 8, and 2 for both $A_{christianity}$ and A_{lgbtq}; in addition, A_{aa} were 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_{lgbtq} depicted similar effort. Similar patterns are also observed in knowledge sharing, structural significance and recruitment.

Interestingly, both A_{news} and A_{ows} exceeded A_{aa} in all 5 dimensions. For instance, the median tweet contribution by A_{news} is 32, 4 times that of A_{aa}. 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. These observations suggest that A_{news} is much better at employing social media mechanisms, perhaps due to the nature of its industry.

While the critical role of progressive news organizations in BlackLivesMatter can be explained by the need of engaging and educating others, what's more surprising is that A_{ows} surpassed both A_{news} and A_{aa} in their commitment to BlackLivesMatter. The median tweet contribution by A_{ows} is 54, the highest amongst all the groups. In fact, as depicted in Figure 2, A_{ows} is the most dedicated amongst all SMIs as measured by number of days in protest and number of protest communities engaged. Consequently, A_{ows} also possess the most favorable network position for recruitment. The median kcore value for A_{ows} is 55 comparing to A_{news}'s 41 and A_{aa}'s 11.

The observation of A_{ows} being an integral recruiter and sustainer for BlackLivesMatter is initially surprising. Further analysis reveals 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 progressed, the protest activity level, measured by tweet count, of A_{ows} waned when that of A_{aa} rose. In fact, as shown in Figure 2(c), A_{aa} is the only SMI that depicted a rise in average participation level [26]. This is consistent with Resource Mobilization[51] which posits that as a movement grows and gains capital, professional movement-specific SMOs will eventually emerge.

4.3.3 Women's Rights. Unlike the extensive collaborative effort shown for BlackLivesMatter, here, we only identify 3 distinct SMIs for Women's Rights: the core SMI *Women*, and peripheral SMIs *Christianity* [27] and *LGBTQ*. Further, we observe that SMOs within *Women* are divided into subgroups focused on distinct issues pertaining to women: domestic abuse, leadership skill for young girls, women's reproductive health, et cetera. None of the subgroups demonstrates substantial difference in their role; thus, for our subsequent analyses, we omit the distinction between them.

We observe that the median protest tweet contribution by all three SMIs is 2, implying that more than half of the SMOs from the core SMI behaved just like the ones from the peripheral SMIs. Nevertheless, we do see more outliers from A_{women}, some of which contributed more than 1000 tweets in the span of a year while none from the other SMIs did. In addition, as shown in Figure 2, we see a modest effort by A_{women} in building communities to facilitate coordination: over 7.5% of A_{women}

[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.

[27] We ensure that SMOs under this SMI are in fact supportive of Women's Rights, such as being Pro-Choice, by manually sample 100 accounts and observe that over 90% of them self identify as progressive religious groups.

engaged with 2 or more protest communities, while less than 2% of $A_{christianity}$ and almost none of A_{lgbtq} did the same. Additionally, the recruitment rate for A_{women} is 5.40 per tweet, approximately 2 and 25 times that of A_{lgbtq} and $A_{christianity}$ respectively. Finally, as shown in Figure 2(d), the average participation intensity of A_{women} demonstrates a downward trend similar to $A_{christianity}$ and A_{lgbtq} as the Women's Rights movement progresses.

As a whole, we show that A_{women} demonstrates comparable commitment level to $A_{christianity}$ and A_{lgbtq} with some outliers being more committed to sustain the movement. Futhermore, A_{women} is more invested in building communities and is more successful with recruitment. The differences are moderate in comparison to BlackLivesMatter, however. One possible explanation is that during the timespan of our dataset, multiple high profile and polarizing events(e.g. Ferguson March) had occurred in BlackLivesMatter, while Women's Rights movement remained in relative abeyance.

4.4 Predictive Modeling of SMOs' Mobilization Volume

Existing social movement literature has used mobilization volume as an informative metric in evaluating SMOs role and importance [45, 56]. Our findings presented in Section 4.2 show that SMOs vary significantly in this dimension. In this section, focusing on BlackLivesMatter, we assess to what extent SMOs' account characteristics, participation pattern, strategic choices, and network structure factors explain the variance in their mobilization volume within online protest communities. Below, we formally define measures related to our modeling task.

Mobilization (MZ): Let S_{blm} be the set of SMOs that participated in BlackLivesMatter. For each $s \in S_{blm}$, we evaluate the mobilization volume using 2 metrics: i) the number of s's followers that participated in the movement *after* the first protest tweet contributed by s, denoted as $MZ^1_{blm,s}$, and ii) recruitment defined in Section 4.2, $MZ^2_{blm,s}$. These two metrics provide subtly different operationalization of mobilization volume. We use them both to ensure our qualitative findings are robust. For both metrics, we fit the following regression model:

$$MZ_{blm,s} = \beta_0 + \underline{\beta} M_{blm,s} + \varepsilon_i \quad (1)$$

where $1 * n$ matrix $M_{blm,s}$ represents the entire set of explanatory variables (e.g. account characteristics of s) and $\underline{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \end{pmatrix}$ is the correlation coefficient matrix.

Account Characteristics: We derive the total number of tweets (irrespective of whether they are related to BlackLivesMatter) posted by s, its account creation timestamp, total numbers of friends and followers as well as the ratio between the two. This set of attributes captures the account's overall activity level on Twitter, its account age, and the scope of connections it has.

Participation Patterns: We determine the total number of protest tweets, the difference in days between the first and last protest tweet contributed by s as well as the ratio between protest tweets and total tweets. Finally, we extract the difference in days between the start of the BlackLivesMatter movement and the first protest tweet contributed by s to capture how early s joined the movement.

SMO to SMO Network Attributes: We derive 3 categories of network structural factors: i) direct ties, ii) prestige, iii) diversity and brokerage.

SMO to SMO Direct Ties: We determine the number of SMOs that are s's friends, followers, or both (follows and is followed by s), as well as the ratio between the number of friends and followers.

Prestige: Curtis and Zurcher [18], and Aveni [4] posit that the more prestigious an SMO's allies are, the more prestigious and effective the SMO itself. Within the network analysis literature, this is analogous to the assertion that the more influential a node's neighbors are, the more important and capable the node itself. This concept is operationalized using the *Pagerank* [12] score. To calculate pagerank, we reuse the directed graphs $G_{blm} = (N_{blm}, E_{blm})$ described in Section 4.1. We then use *SNAP* [44], a popular network analysis tool, to derive the pagerank score for s.

Diversity and Brokerage: Aveni [4] theorizes that the more diverse an SMO's linkages to other organizations, the more different resources it can utilize to sustain itself and strive for its objectives. This concept is similar to previous studies on network brokerage and *betweenness centrality*. Here, we also make a distinction between local and global betweenness centrality. Local betweenness centrality, or *local transitivity*, measures the fraction of s's allies that are allies themselves. Prior work [16] suggests that ideas proposed by those with high local transitivity are more likely to be viewed as novel and valuable. In contrast, global betweenness centrality of s measures the number of shortest path between all pairs of S_{blm} that go through s [31]. We use *iGraph* to derive local transitivity and the *SNAP* global betweenness centrality of s.

Strategic Choices: we broadly partition SMO strategies into 4 categories: knowledge sharing, appeal to emotion, direct lobbying, and media effort.

Knowledge Sharing: SMOs have a much higher usage of hyperlinks and hashtags than the baseline Twitter average for the purpose of sharing information beyond the 140 character limitation set by Twitter [24, 34]. We calculate the fraction of protest tweets contributed by s that contains at least 1 hyperlink, and the average number of hashtags per tweet.

Appeal to Emotion: Related work suggests that political tweets with negative emotional sentiment are more likely to be shared [65]. Here, we first manually label a subset of protest tweets as *positive*, *negative*, or *neutral* in sentiment. [28] Next, for the labeled subset, We apply sentiment scoring methods VADAR [37] and SentiEval [43] to extract relevant text features from each protest tweet and build a supervised multiclass learner; it has 66% accuracy which is better than using each method separately. [29] We classify the remaining unlabeled protest tweets by S_{blm} as *positive*, *negative*, or *neutral*. We observe that the vast majority of tweets, 60%, are neutral, with 32.3% of all tweets negative, and 0.7% positive. Finally, we derive the fraction of protest tweets by s that connotes positive sentiment, and the fraction with negative sentiment.

Direct Lobbying: This strategy is used by s when it directly approaches or interacts with lawmakers using social media to lobby for its cause [34]. Here, we identity Twitter accounts that belong to federal legislators by using existing dataset by Tauberer [67]. To determine accounts that belong to state, city, or county level elected officials (e.g. state senator, county clerk, et cetera) we first aggregate a list of elected positions of (i.e. job titles)[30], we then derive a list of accounts that contain at least one of the aggregated occupation titles within the user profile description field. We obtain approximately 3K such accounts. Next, we randomly select 300 accounts and examine whether each account indeed belongs to an elected official. We observe that 224 or 79% of the select subset correctly match to actual officials. We calculate the number of protest tweets by s that mentions or is a retweet of these accounts.

Media Effort: Prior studies observe that SMOs actively engage with newspapers and journalists to bring coverage and push public agenda aligned with their goals [2]. We identity 4.4K news [31] and

[28]We randomly select 400 protest tweets each contributed by a distinct SMO. Of the 400 tweets, 60% (or 240) is labeled as *neutral*, 32.2% is *negative*, and the remaining *positive*.

[29]We select features based on [43]. We then use Linear SVM with parameters tuned by GridSearchCV and 10-fold cross validation using the *scikit-learn* library. Performance of 2-class classification is comparable at 70% accuracy. In addition, increasing labeled set size does not contribute to performance: when trained and tested on a labeled set of size 100, the classifier has an accuracy of 63%. Accuracy is 65% for a set of size 200. We observe no change from size 300 to 400.

[30]We combine data from multiple different websites: *wa.gov*, *usa.gov*, and *politicalcampaigningtips.com*.

[31]We use BING to crawl for Twitter accounts of all websites, approximately 20K, listed under the *News* category on Alexa and DMOZ. For instance, given the site *www.cnn.com*, we search BING using the query *"www.cnn.com twitter account"*, and extract the first Twitter account returned. We randomly select and check 300 websites and their corresponding crawled accounts, 224 or 75% of the results are correct. Out of the remaining, 49 or 16.3% of the sample do not have a Twitter account, 6 or 2% participated in one of the movements but the crawler returned the wrong account. Finally, 21 or 7% of the sampled sites were matched to the wrong account and also did not participate in either movement.

50.6K individual reporter[32] Twitter accounts in our dataset. We determine the number of protest tweets by s that mentions or is a retweet of these accounts in the BlackLivesMatter movement.

Model Performance: Using Equation 1, We fit and evaluate 5 separate models; 4 of which are partial models using only predictors from a single category: i) account characteristics, ii) participation patterns, iii) network attributes, iv) strategic factors, and 1 complete model containing all predictors. In addition, we assess each model using 2 distinct sets of SMOs: i) SMOs from the African American SMI, denoted as A_{aa}, and ii) SMOs from all SMIs, A_{all}. We make this distinction to differentiate SMOs belonging to core SMI from SMOs of more peripheral SMIs.

The adjusted R^2 values for all models estimated through cross validation[33] are shown in Table 6. We see that the complete MZ^1_{blm} models for A_{aa} and A_{all} have adjusted R^2 of 0.95 and 0.93 respectively; that is, we are able to explain a substantial fraction of the variance in SMOs volume with respect to MZ^1_{blm}. Furthermore, account characteristics and network attributes are much more predictive of SMOs volume comparing to participation patterns and strategies. In fact, the later two categories have adjusted R^2 of only 0.08 and 0.064 respectively. In addition, while network attributes, accounting for 92% of all variance, dominated all the other categories of predictors for A_{all}, account characteristics are as equally predictive as network attributes for A_{aa}, In fact, account characteristics alone are able to explain 81.9% of all variance in A_{aa}. This is in line with prior observations under Section 4.3: our dataset contains the early period of the BlackLivesMatter movement during which movement specific SMO accounts have yet to build network connections with other SMOs, making the individual account characteristics of A_{aa} as indicative of their volume as network attributes. For MZ^2_{blm}, our models have a more modest predictive power for A_{aa} and A_{all}: R^2 of 0.50, and 0.52 respectively. Additionally, we also see that variables in the participation patterns category, performs better at predicting overall volume in MZ^2_{blm}. One intuitive explanation is that for a follower of s, it matters little when or how intensely s participates given that s's tweets show up on a follower's feeds regardless; but, for someone who does not follow s, that person's exposure to s is likely dependent on s's participation patterns such as how early s joined the movement . However, the significance of network attributes is consistent across the two definitions.

Coefficients estimates of each predictor for the complete MZ^1_{blm} model using A_{aa} are plotted in Figure 3. As depicted, within the account characteristics category, the coefficients for the number of followers and friends are 1070.8 and -308.1 respectively (i.e. one standard deviation change in follower count is correlated with an additional 1.1K followers being mobilized), implying that having more followers and fewer friends are positively correlated with SMOs being more successful at mobilization. For network attributes, we observe that having higher betweenness centrality is positively associated with greater mobilization volume, implying that SMOs with increased "brokerage" or having access to more varied information, are able to better mobilize their followers. In addition, the number of bidirectional ties with other SMOs are also positively correlated with mobilization volume. Finally, SMOs that participated earlier in movements and/or have a higher engagement with news or reporter accounts also have a higher mobilization volume.

5 DISCUSSION & LIMITATIONS

In this paper, we provided the first automated method in classifying social movement organizations (SMOs) at scale. Using this method, we identified over 50K SMOs participating in two significant social media protests: BlackLivesMatter and Women's Rights. We conducted in-depth comparisons of SMOs' involvement in online protests to that of other organizations (OTHOs) and individuals,

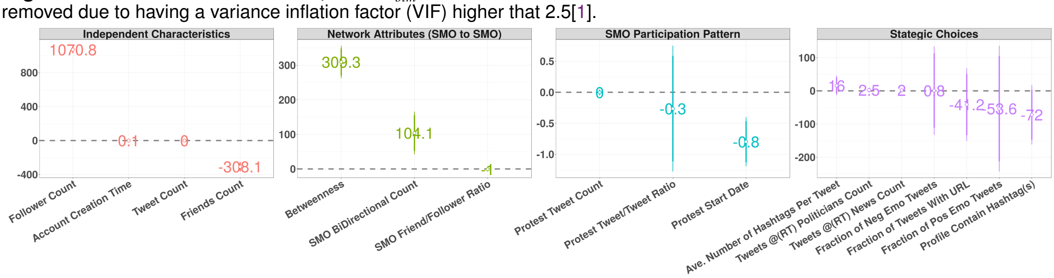
[32] We first aggregate a list of news media occupation such as *reporter* or *journalist*, we then filter and select the subset of users that listed any one of the relevant occupations in the profile description field.

[33] We first assign data into 10 bins using the value of MZ_{blm}; we then split data using the bin number into 10 stratified folds (training and test datasets). Finally, we use reassess R^2 for out of sample test data.

Table 6. Adjusted R^2 for all Mobilization Models.

Metric	Dataset	R^2. Account Characteristics	R^2.Network Attributes	R^2.Participation Pattern	R^2.Strategies	R^2.Complete
MZ^1_{blm}	African American SMI	0.819	0.896	0.08	0.064	0.956
	All SMIs	0.293	0.919	0.164	0.104	0.928
MZ^2_{blm}	African American SMI	0.137	0.342	0.313	0.151	0.495
	All SMIs	0.178	0.37	0.339	0.195	0.519

Fig. 3. Coefficients Plot of the Complete MZ^1_{blm} Model for SMOs in the African American SMI. Note*. Some variables were removed due to having a variance inflation factor (VIF) higher that 2.5[1].



and demonstrated that beyond the eye-catcher SMOs which are highly visible to the movements, an average SMO is very similar to an average individual both in terms of behavior and significance. We also observed 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. Finally, we proposed models that are highly predictive of SMOs mobilization volume.

There are several limitations and future directions worth noting. First, our analyses rely on Twitter data and do not address SMOs' characteristics and role outside of the platform. Second, we focus on two successful and high profile movements. Future work that extends this understanding to other—less popular and successful—movements is necessary to broaden our understanding. Third, our method in categorizing SMOs into SMIs is unable to assign a fraction of SMOs into coherent SMIs. Fourth, future work should distinguish individuals who are members of SMOs from non-members to further explore SMOs induced impact on social media activism.

Lastly, we discuss several important implications of our study. First, considering the extensive involvement of SMOs from peripheral SMIs in BlackLivesMatter, 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. Next, 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. In addition, our model highlighted the importance of network structural attributes in relation to SMOs mobilization volume which potentially have real world applications for organizational leaders and platform owners. Based on our observations, SMOs can possibly improve their mobilization volume by focusing on strategically building connections with others SMOs from a more diverse set of SMIs. In fact, the SMOs dataset we gathered can serve as valuable candidates. Finally, platform owners can also explore the SMOs network and build a recommendations system for SMOs to make additional organizational connections.

REFERENCES

[1] Paul Allison. 2012. When Can You Safely Ignore Multicollinearity? https://statisticalhorizons.com/multicollinearity
[2] Kenneth T Andrews and Neal Caren. 2010. Making the news: Movement organizations, media attention, and the public agenda. *American Sociological Review* 75, 6 (2010), 841–866.
[3] 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.
[4] Adrian F. Aveni. 1978. Organizational Linkages and Resource Mobilization: The Significance of Linkage Strength and Breadth. *The Sociological Quarterly* 19, 2 (1978), 185–202. http://www.jstor.org/stable/4105631
[5] Pablo Barberá, 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 (2015), e0143611.
[6] Mathieu Bastian, Sebastien Heymann, Mathieu Jacomy, et al. 2009. Gephi: an open source software for exploring and manipulating networks. *Icwsm* 8 (2009), 361–362.
[7] W Lance Bennett. 2012. The personalization of politics: Political identity, social media, and changing patterns of participation. *The ANNALS of the American Academy of Political and Social Science* 644, 1 (2012), 20–39.
[8] Bruce Bimber, Andrew Flanagin, and Cynthia Stohl. 2012. *Collective action in organizations: Interaction and engagement in an era of technological change*. Cambridge University Press.
[9] 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.
[10] Phillip Bonacich. 1987. Power and centrality: A family of measures. *American journal of sociology* 92, 5 (1987), 1170–1182.
[11] 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.
[12] Sergey Brin and Lawrence Page. 2012. Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer networks* 56, 18 (2012), 3825–3833.
[13] Jon Bruner. 2013. Tweets Loud and Quiet. https://www.oreilly.com/ideas/tweets-loud-and-quiet
[14] 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.
[15] Ceren Budak and Duncan J Watts. 2015. Dissecting the Spirit of Gezi: Influence vs. Selection in the Occupy Gezi Movement. *Sociological Science* 2 (2015), 370–397.
[16] Ronald S Burt. 2009. *Structural holes: The social structure of competition*. Harvard university press.
[17] 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.
[18] Russell L. Curtis and Louis A. Zurcher. 1973. Stable Resources of Protest Movements: The Multi-Organizational Field. *Social Forces* 52, 1 (1973), 53–61. https://doi.org/10.2307/2576423
[19] Munmun De Choudhury, Nicholas Diakopoulos, and Mor Naaman. 2012. Unfolding the event landscape on twitter: classification and exploration of user categories. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, 241–244.
[20] Munmun De Choudhury, Shagun Jhaver, Benjamin Sugar, and Ingmar Weber. 2016. Social Media Participation in an Activist Movement for Racial Equality.. In *ICWSM*. 92–101.
[21] Nadbor Drozd. 2016. Text Classification With Word2Vec. http://nadbordrozd.github.io/blog/2016/05/20/text-classification-with-word2vec/
[22] 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
[23] Jennifer Earl and Katrina Kimport. 2011. *Digitally enabled social change: Activism in the Internet age*. Mit Press.
[24] Heather R Edwards and Richard Hoefer. 2010. Are social work advocacy groups using Web 2.0 effectively? *Journal of Policy Practice* 9, 3-4 (2010), 220–239.
[25] 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
[26] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. The rise of social bots. *Commun. ACM* 59, 7 (2016), 96–104.
[27] Tina Fetner and Brayden G King. 2016. Three-Layer Movements, Resources, and the Tea Party. *Understanding the Tea Party Movement* (2016), 35.
[28] Ronen Fluss, David Faraggi, and Benjamin Reiser. 2005. Estimation of the Youden Index and its associated cutoff point. *Biometrical journal* 47, 4 (2005), 458–472.

[29] Michelle C Forelle, Philip N Howard, Andrés Monroy-Hernández, and Saiph Savage. 2015. Political bots and the manipulation of public opinion in Venezuela. (2015).

[30] Deen Freelon, Charlton D McIlwain, and Meredith D Clark. 2016. Beyond the hashtags:# Ferguson,# Blacklivesmatter, and the online struggle for offline justice. (2016).

[31] Linton C Freeman. 1977. A set of measures of centrality based on betweenness. *Sociometry* (1977), 35–41.

[32] 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).

[33] 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).

[34] 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.

[35] Stefanie Haustein, Timothy D Bowman, Kim Holmberg, Andrew Tsou, Cassidy R Sugimoto, and Vincent Larivière. 2016. Tweets as impact indicators: Examining the implications of automated âĂIJbotâĂİ accounts on Twitter. *Journal of the Association for Information Science and Technology* 67, 1 (2016), 232–238.

[36] 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

[37] Clayton J Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media*.

[38] Sarah J Jackson and Brooke Foucault Welles. 2015. Hijacking# myNYPD: Social media dissent and networked counterpublics. *Journal of Communication* 65, 6 (2015), 932–952.

[39] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2016. Bag of Tricks for Efficient Text Classification. *arXiv preprint arXiv:1607.01759* (2016).

[40] David Karpf. 2012. *The MoveOn effect: The unexpected transformation of American political advocacy*. Oxford University Press.

[41] Darius Kazemi. 2015. Occupations. https://github.com/dariusk/corpora/blob/master/data/humans/occupations.json

[42] Sunghwan Mac Kim, Cecile Paris, Robert Power, and Stephen Wan. 2017. Distinguishing Individuals from Organisations on Twitter. In *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 805–806.

[43] Svetlana Kiritchenko, Xiaodan Zhu, and Saif M Mohammad. 2014. Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research* 50 (2014), 723–762.

[44] Jure Leskovec and Rok Sosič. 2016. SNAP: A General-Purpose Network Analysis and Graph-Mining Library. *ACM Transactions on Intelligent Systems and Technology (TIST)* 8, 1 (2016), 1.

[45] 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.

[46] Fiona Martin and Mark Johnson. 2015. More efficient topic modelling through a noun only approach. In *Australasian Language Technology Association Workshop 2015*. 111.

[47] Winter Mason and Siddharth Suri. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods* 44, 1 (2012), 1–23.

[48] 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

[49] 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/"

[50] 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

[51] 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

[52] James McCorriston, David Jurgens, and Derek Ruths. 2015. Organizations Are Users Too: Characterizing and Detecting the Presence of Organizations on Twitter.. In *ICWSM*. 650–653.

[53] 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.

[54] 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

[55] 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.

[56] 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.
[57] Alexandra Olteanu, Ingmar Weber, and Daniel Gatica-Perez. 2015. Characterizing the demographics behind the# blacklivesmatter movement. *arXiv preprint arXiv:1512.05671* (2015).
[58] Marco Pennacchiotti and Ana-Maria Popescu. 2011. Democrats, republicans and starbucks afficionados: user classification in twitter. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 430–438.
[59] 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.
[60] 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.
[61] Clay Shirky. 2008. *Here comes everybody: The power of organizing without organizations*. Penguin.
[62] 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
[63] 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
[64] Kate Starbird and Leysia Palen. 2012. (How) will the revolution be retweeted?: information diffusion and the 2011 Egyptian uprising. In *Proceedings of the acm 2012 conference on computer supported cooperative work*. ACM, 7–16.
[65] Stefan Stieglitz and Linh Dang-Xuan. 2013. Emotions and information diffusion in social mediaâĂŤsentiment of microblogs and sharing behavior. *Journal of Management Information Systems* 29, 4 (2013), 217–248.
[66] J Tang, S Alelyani, and H Liu. 2014. Data classification: algorithms and applications. *Data Mining and Knowledge Discovery Series, CRC Press* (2014), 37–64.
[67] Joshua Tauberer. 2017. Members of the United States Congress, 1789-Present. https://github.com/unitedstates/congress-legislators
[68] 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
[69] Yannis Theocharis, Will Lowe, Jan W. van Deth, and Gema GarcÃ■a-Albacete. 2015. Using Twitter to mobilize protest action: online mobilization patterns and action repertoires in the Occupy Wall Street, Indignados, and Aganaktismenoi movements. *Information, Communication & Society* 18, 2 (Feb. 2015), 202–220. https://doi.org/10.1080/1369118X.2014.948035
[70] Robert L Thorndike. 1953. Who belongs in the family? *Psychometrika* 18, 4 (1953), 267–276.
[71] Onur Varol, Emilio Ferrara, Clayton A Davis, Filippo Menczer, and Alessandro Flammini. 2017. Online human-bot interactions: Detection, estimation, and characterization. *arXiv preprint arXiv:1703.03107* (2017).
[72] Ingmar Weber, Venkata R Kiran Garimella, and Alaa Batayneh. 2013. Secular vs. islamist polarization in Egypt on Twitter. In *Advances in Social Networks Analysis and Mining 2013, ASONAM '13*. ACM, 290–297.
[73] 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
[74] 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
[75] Jinhong Zhong, Ke Tang, and Zhi-Hua Zhou. 2015. Active Learning from Crowds with Unsure Option.. In *IJCAI*. 1061–1068.