Understanding and Identifying Advocates for Political Campaigns on Social Media

Suhas Ranganath†, Xia Hu⋆, Jiliang Tang‡, Huan Liu†
† Arizona State University, ⋆ Texas A&M University, ‡ Yahoo! Labs
{srangan8,huan.liu}@asu.edu, hu@cse.tamu.edu, jlt@yahoo-inc.com

ABSTRACT

Social media is increasingly being used to access and disseminate information on sociopolitical issues like gun rights and general elections. The popularity and openness of social media makes it conducive for some individuals, known as advocates, who use social media to push their agendas on these issues strategically. Identifying these advocates will caution social media users before reading their information and also enable campaign managers to identify advocates for their digital political campaigns. A significant challenge in identifying advocates is that they employ nuanced strategies to shape user opinion and increase the spread of their messages, making it difficult to distinguish them from random users posting on the campaign. In this paper, we draw from social movement theories and design a quantitative framework to study the nuanced message strategies, propagation strategies, and community structure adopted by advocates for political campaigns in social media. Based on observations of their social media activities manifesting from these strategies, we investigate how to model these strategies for identifying them. We evaluate the framework using two datasets from Twitter, and our experiments demonstrate its effectiveness in identifying advocates for political campaigns with ramifications of this work directed towards assisting users as they navigate through social media spaces.

Keywords

Advocacy; Political Campaigns; User Interactions

1. INTRODUCTION

Social media is emerging to be a popular information channel for sociopolitical issues of broad importance for e.g. elections and gun rights. It provides access to a wide range of perspectives on these issues, enabling users to form independent opinions. Owing to this, millions of people are using social media to seek information, gather perspectives, and participate in discussions on these important issues. The popularity of social media as an information channel for important issues has given rise to individuals who use it to try and push their agenda for political campaigns [17]. Media advocacy is defined in the the literature as "the strategic use of mass media to advance a social or public initiative" [19]. We are motivated by this to define advocates for political campaigns on social media as individuals who use social media to advance strategically their agenda for a given political campaign. During the 2014 Indian elections, for example, a set of individuals formed an organization called NaMo Brigade with the motto of "Mission: Narendra Modi as PM" and used social media platforms to advocate for the election of Narendra Modi as Prime Minister [8].

Although these campaigns have considerable social media presence [37], it is difficult to identify individual accounts of advocates. Designing algorithms to identify accounts of individual advocates can better inform users as they navigate through social media spaces. People accessing information about an issue, for e.g. an election, through social media can be notified whether a given account is an advocate before reading their messages. On the other hand, understanding and modeling the characteristics of social media advocates can help campaign managers in recruiting new advocates for their digital political campaigns.

This task faces several challenges. First, advocates employ nuanced message construction and propagation strategies to shape user opinion and increase the spread of their messages, making it difficult to distinguish them from random users posting on issues related to the campaign. Second, these strategies are very diverse, manifesting both in the activity patterns restricted to individual advocates like constructing persuasive messages, and multiple relational patterns like shared language and interactions, making it a challenge to study collectively them in a unified model.

Theoretical constructs of strategies for message construction, propagation, and community formation by advocates have been extensively studied in social sciences. Social movement theory records persuasive language and high degrees of emotion in the messages of advocates in their attempts to shape the opinion of people [28]. The literature on campaign communications [13] studies the widespread use of focused messaging for effective communication during political campaigns. Also, distinctive language patterns shared among people with similar affiliations foster easier communication of messages between them [32]. To increase the reach of messages during political campaigns, the utility of popular users for widespread propagation has been studied in [15]. Formation of social and interaction networks between advocates for

WSDM 2016 February 22-25, 2016, San Francisco, CA, USA
© 2016 ACM. ISBN 978-1-4503-3716-8/16/02…$15.00
DOI: http://dx.doi.org/10.1145/2835776.2835807

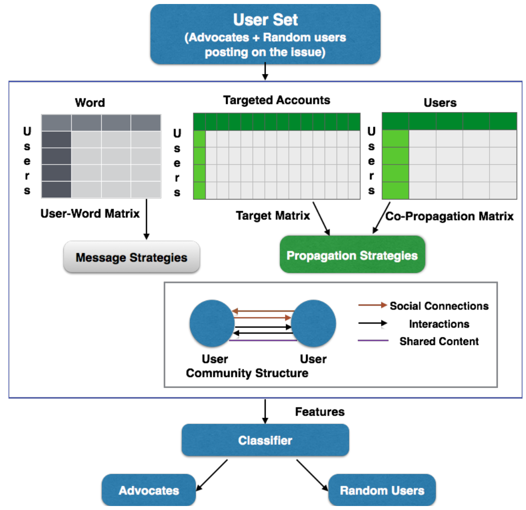


Figure 1: The proposed framework to identify advocates for political campaigns on social media

the purpose of easier coordination and communication has been investigated in social movement literature [28].

In this paper, we model the nuanced message strategies, propagation strategies, and community structure of advocates guided by sociological literature and integrate them to identify advocates for political campaigns on social media. We primarily focus on the following questions: How to model the nuanced strategies of advocates for political campaigns on social media? How to integrate them to design a unified framework for identifying advocates for political campaigns in social media?

The primary contributions of this work are:

- A definition of the problem of identifying advocates for political campaigns on social media,
- A computational framework to gain insights into the strategies of the advocates and identify them by collectively modeling their strategies, and
- Evaluation of the framework in identifying advocates of political campaigns in the social media platform Twitter using two real-world datasets.

2. PROBLEM STATEMENT

In this section, we introduce notations and terms used in the paper and formally define the problem statement. We first define some notations. Boldface uppercase letters (e.g $\mathbf{X}$) denote matrices and boldface lowercase letters (e.g. $\mathbf{x}$) vectors. $\mathbf{X}_{\mathrm{ij}}$ signifies the element in the i^{th} row and j^{th} column of $\mathbf{X}$ and the i^{th} column of $\mathbf{X}$ by $\mathbf{X}_i$. The Frebonius norm of $\mathbf{X}$ is denoted as $||\mathbf{X}||_{\mathrm{F}} = \sqrt{\sum_{\mathrm{i,j}} \mathbf{X}_{\mathrm{ij}}^2}$. The n mode vector product of a tensor $\mathcal{M} \in \mathbb{R}^{D_1 \times D_2 \times .. \times D_N}$ with $\mathbf{x} \in \mathbb{R}^{D_n}$ is given by $\mathcal{M} \times_n \mathbf{x}$ and results in a tensor of size $D_1 \times D_2 \times .. \times D_{n-1} \times D_{n+1} \times ... \times D_N$ whose each element is given by $(\mathcal{M} \times_n \mathbf{x})_{d_1 d_2 .. d_{n-1} d_{n+1} ... d_N} = \sum_{d_n=1}^{D_n} \mathcal{M}_{d_1 d_2 .. d_N} \mathbf{x}_{d_n}$. The set of advocates is denoted as $\mathbf{a}$ and the set of random users using keywords related to the issue as $\mathbf{v}$. The set of users is denoted as $\mathbf{u} = [\mathbf{a}, \mathbf{v}]$ and the total number of users as N.

We now define terms related to different kinds of strategies of advocates that can be characterized. Message strategies deal with the construction of status messages of advocates with the aim of shaping the opinions of people. We develop characterizations on four types of message strategies possibly present in the status messages of advocates: persuasive language, a high degree of emotion, topical focus and shared language patterns. We construct the user word matrix $\mathbf{S} \in \mathbb{R}^{\mathrm{N} \times \mathrm{l}}$ from the status messages with tf-idf weighting, where l is the total number of words. Shared language patterns between two users in $\mathbf{u}$ are modeled taking hashtags as instances of language patterns. For each user $\mathbf{u}_{\mathrm{i}} \in \mathbf{u}$, we construct a vector of hashtags $\mathbf{h}_{\mathbf{u}_{\mathrm{i}}}$ from his status messages. We define matrices capturing shared hashtag information as $\mathcal{Z}^1$, where $\mathcal{Z}^1_{\mathrm{ij}} = \mathrm{jac\_sim}(\mathbf{h}_{\mathbf{u}_{\mathrm{i}}}, \mathbf{h}_{\mathbf{u}_{\mathrm{j}}})$. jac_sim($\mathbf{x}$,$\mathbf{y}$) indicates the Jaccard similarity between $\mathbf{x}$ and $\mathbf{y}$.

Propagation strategies comprise of the strategies employed by advocates to increase the spread of their messages. We model the propagation strategies of advocates from their targeting and co-propagation behavior. For each user in $\mathbf{u}_{\mathrm{i}} \in \mathbf{u}$, let $\mathbf{ta}_{\mathrm{u}}$ denote the set of people targeted by the user. We define the targeting matrix $\mathbf{T} \in \mathbb{R}^{\mathrm{N} \times \mathrm{R}}$ where $\mathbf{T}_{\mathrm{ij}}$ is equal to the number of times $\mathbf{u}_{\mathrm{i}}$ has targeted $\mathbf{r}_{\mathrm{j}}$ where $\mathbf{r} = \bigcup_{\mathrm{i} \in \mathbf{u}} \mathbf{ta}_{\mathrm{i}}$ and R is the total number of users in $\mathbf{r}$. We next define the co-propagation network as $\mathbf{P} \in \mathbb{R}^{\mathrm{N} \times \mathrm{N}}$, where $\mathbf{P}_{\mathrm{ij}}$ is the number of times $\mathbf{u}_{\mathrm{j}}$ has propagated a message of $\mathbf{u}_{\mathrm{i}}$.

Community structure deals with the patterns arising from the networks formed by advocates to facilitate easier communication and coordination. We model the community structures arising from social connections and interactions between advocates. The social connection matrix is defined as $\mathcal{Z}^2 \in \mathbb{R}^{\mathrm{N} \times \mathrm{N}}$. The value of $\mathcal{Z}^2_{\mathrm{ij}}$ is 1 if $\mathbf{u}_{\mathrm{i}}$ connects to $\mathbf{u}_{\mathrm{j}}$ and 0 otherwise. We also capture the interactions between the users in $\mathcal{Z}^3 \in \mathbb{R}^{\mathrm{N} \times \mathrm{N}}$ where $\mathcal{Z}^3_{\mathrm{ij}}$ is equal to the number of times $\mathbf{u}_{\mathrm{i}}$ interacts with $\mathbf{u}_{\mathrm{j}}$. In addition, we define the tensor $\mathcal{Z}$ to hold the information contained in the relational matrices $\mathcal{Z}^1, \mathcal{Z}^2$ and $\mathcal{Z}^3$, where $\mathcal{Z}_{\mathrm{ijt}} = \mathcal{Z}^{\mathrm{t}}_{\mathrm{ij}}$.

We model the message strategies, propagation strategies, and community structure and study them in a unified supervised learning framework to identify advocates for political campaigns on social media. The problem statement can then be stated as follows: *"Given a set of advocates for a given political campaign on social media* $\mathbf{a}$ *and a set of random users posting on the campaign* $\mathbf{v}$*, their status message matrix* $\boldsymbol{S}$*, the targeting matrix* $\boldsymbol{T}$*, the propagation network* $\boldsymbol{P}$*, and the relational tensor* $\mathcal{Z}$*, determine if a new user x is an advocate of the political campaign."*

3. QUANTIFYING STRATEGIES

In this section, we study the different nuanced strategies employed by advocates for political campaigns on social media drawing from theoretical constructs present in sociological literature and present ways to model them. We first study the employed strategies in terms of their message strategies, propagation strategies, and community structure. We explore each of them in detail, and then present ways to model them to derive characteristics possibly capable of distinguishing between advocates for political campaigns and random users posting on the campaign.

3.1 Quantifying Message Strategies

Message strategies deal with patterns from the construction of the status messages of advocates. Advocates can employ persuasive language in their status messages and attempt to sway opinions of other users. Parallels can be seen in the sociological literature which documents the high use of persuasive language during social movements [28]. The status messages of advocates can also contain a high degree of emotions both positive; when they try to generate positive feelings about their campaign, or negative; when they try to generate feelings of anger, fear, and anxiety [28].

Message strategies of advocates can also manifest across a set of status messages. For instance, advocates can concentrate their status messages around a small number of topics in their attempts at effective communication, resulting in high topical focus. Parallels can be seen in studies of campaign communications [13], which shows a widespread use of focused messaging. An advocate can also share distinctive language patterns with other advocates to support common issues and facilitate easier communication. Shared language patterns among people with similar affiliations have been shown in speech codes theory to foster easier communication [32]. Next, we present ways to model characteristics arising from persuasive language, emotions, focused messaging, and shared language patterns.

The use of persuasive language can be quantified by modeling theoretical principles of persuasion [11]. We consider two principles of reason and affinity [11] and model their occurrence. People employ reason as a rational justification for their views while persuading others. The number of words related to reason [14] in the status messages of a user is used to quantify reason. Expressions of affinity can also be used as a tool for persuasion by using words conveying liking, compliments, and association. We use social words in the LIWC corpus [31] to model expressions of affinity and count their occurrence in the messages of each user. We postulate that advocates in $\mathbf{a}$ use a higher number of words denoting persuasion than random users $\mathbf{v}$ posting on the campaign.

The emotional content in the posts of the users in $\mathbf{u}$ can be modeled using the positive and negative emotional words from the LIWC corpus [31]. We postulate that advocates in social media use a higher number of emotional words, both positive and negative than random users posting on an issue. A higher use of emotion can be an indication that advocates have a strong belief in their cause, which separates them from paid workers posting promotional comments for a campaign, who display fewer emotions in their posts [24].

To model the topical focus of a user in $\mathbf{u}$, we first compute the topic distribution using LDA [6] on the user word matrix $\mathbf{S}$. The topic model results in the user topic matrix $\mathbf{DT} \in \mathbb{R}^{\mathrm{n}\times\mathrm{t}}$, where $\mathbf{DT}_{\mathrm{ij}}$ is the number of times a word of user $\mathbf{u}_{\mathrm{i}}$ has been assigned to topic $\mathbf{t}_{\mathrm{j}}$ and t is the number of topics. To obtain the topic probability distribution, the document-topic matrix can be normalized $\mathbf{DT}$ to obtain $\mathbf{DT}'$, where each row of $\mathbf{DT}'$ contains the probability distribution over topics of a user. The entropy of the topic distribution from the corresponding row of $\mathbf{DT}$ for each user i to construct a vector $\mathbf{e}_{\mathrm{i}}$ where $\mathbf{e}_{\mathrm{i}} = \sum_{\mathrm{j=1}}^{\mathrm{j=t}} -\mathbf{DT}'_{\mathrm{ij}}\log(\mathbf{DT}'_{\mathrm{ij}})$. It is evident that a lower value of entropy for a user implies greater concentration on fewer topics in his messages and a higher value implies distribution over a larger number of topics. Therefore, the higher the topical focus in the status messages of a user, the lower the value of his entropy. We postulate that advocates have a higher topical focus in their messages than random users posting on the campaign.

We next model the shared language patterns among users in $\mathbf{u}$ and use hashtags as instances of language patterns. For each advocate $\mathrm{a} \in \mathbf{a}$, let $\mathcal{Z}^1_{\mathrm{ab}}$ denote the amount of hashtags he shares with any other advocate $b \neq a \in \mathbf{a}$ measured by Jaccard similarity. Similarly, let $\mathcal{Z}^1_{\mathrm{av}}$, where $\mathrm{v} \in \mathbf{v}$ is a random user posting on issues related to the campaign. We postulate that an advocate for a political campaign shares a higher amount of hashtags with other advocates than with random users in $\mathbf{v}$ posting on the campaign. We evaluate these characteristics using the datasets in Section 4.2.1.

Until now, we characterized and modeled message strategies capable of distinguishing advocates for political campaigns from random users posting on the issue like persuasive language, emotion, topical focus and shared language patterns with other advocates. We next examine characteristics of propagation strategies employed by advocates to increase the reach of these messages.

3.2 Quantifying Propagation Strategies

We examine the propagation strategies of advocates for political campaigns on social media, focusing on their targeting and co-propagation behavior. Social media enables advocates to target specific users for propagating information. We propose that advocates target popular users more frequently than random users posting on the campaign as popular users help to get messages across to a wider audience [15]. We then investigate how advocates assist each other in spreading their messages.

We first model the targeting behavior of the users in $\mathbf{u}$. Let $\mathbf{r}$ is the set of people targeted by all users in $\mathbf{u}$ as defined in Section 2. Taking the number of users connecting to a user as a measure of his popularity, we construct the vector $\mathbf{c}$ where $\mathbf{c}_{\mathrm{i}}$ is the number of people connecting to $\mathbf{r}_{\mathrm{i}}$. We postulate that the attention of advocates is more skewed towards users with higher popularity than the attention of random users posting on the campaign. To model this postulate, we compute the vector $\mathbf{sta} = \mathbf{Tc}$, where $\mathbf{sta}_{\mathrm{k}}$ is the sum of number of times a user k targets an user $\mathbf{r}_{\mathrm{i}}$ weighted by $\mathbf{c}_{\mathrm{i}}$, the number of users connecting to $\mathbf{ta}_{\mathrm{i}}$. The value of $\mathbf{sta}_{\mathrm{k}}$ is higher if the user k targets popular users a higher number of times. Therefore, our postulate is satisfied when advocates for a political campaign have a higher value of $\mathbf{sta}$ than random users posting on the campaign.

We next model the co-propagation behavior of users in $\mathbf{u}$. Advocates will be more interested in propagating messages of other advocates, and also, their messages will be more likely to be propagated by other advocates than random users posting on the campaign. Based on this, we characterize advocates by their hubs and authority scores [22] in the information propagation network $\mathbf{P}$. We compute hubs and authority scores of users in $\mathbf{u}$ using the information propagation network $\mathbf{P}$ and postulate that advocates have higher hub and authority scores than random users posting on the campaign. We evaluate these characteristics using the datasets in Section 4.2.2.

Until now, we proposed characteristics of advocates for political campaigns on social media from their message strategies and propagation strategies along with methods for modeling them. We next propose characteristics related to the community structure arising from their relationships with other advocates for the campaign.

3.3 Quantifying Community Structure

Community structure deals with the patterns arising from the networks formed by advocates to facilitate easier communication and coordination. Social media provides opportunities for advocates to connect to each other through many different types of relationships. Advocates can form social connections, interact with each other for coordination, and carry out conversations. The formation of networks of social connections and interactions between advocates for communication and coordination have been studied in theoretical studies of social movements [28]. Social connections and interactions between advocates can give rise to the similarity in community memberships. We now postulate a few underlying hypothesis to establish the similarity in community memberships between advocates in social media

We first present postulates underlying community structure arising from social connections and interactions of advocates. For each advocate $a \in \mathbf{a}$, let $\mathcal{Z}^2_{ab}$ be 1 if a connects to b and 0 otherwise, where b is another advocate $b \neq a \in \mathbf{a}$ and $\mathcal{Z}^2$ is defined in Section 2. Similarly, let $\mathcal{Z}^2_{av}$ be 1 if a connects to b and 0 otherwise, where $v \in \mathbf{v}$ is a random user posting on issues related to the campaign We then postulate that advocates are more likely to form social connections with other advocates than with random users posting on the campaign. We follow a similar procedure using $\mathcal{Z}^3$ to postulate that advocates are more likely to interact with each other than with random users posting on the campaign. These postulates underly that advocates have similar community memberships for different types of relationships. Are these community memberships of advocates are similar when measured across relationship types?

To model this, we first select the users with whom the advocates have at least one type of relation with. For each advocate, we construct a vector $\mathbf{co}_{ab}$, $b \neq a \in \mathbf{a}$, where each element is the number of types a pair of advocates have relations in. Similarly, we construct the vector $\mathbf{co}_{av}$, $v \in \mathbf{v}$, where each element is the number of types of relations between a pair of advocate and a random user posting on the campaign. We postulate that given an advocate has established one type of relation with a user, he has a significantly higher propensity to form more types of relation if the user is another advocate than if he is a random user posting on the campaign. The postulate, if verified, underlies that community memberships of advocates are shared across different relationship types, and hence, they can be jointly inferred by combining different relationship types in an efficient manner.

In this section, we draw from theoretical constructs in sociological literature to propose different characterizations of the nuanced message strategies, propagation strategies, and community structure of advocates for political campaigns on social media. Next, we are going to use two real-world datasets from Twitter to evaluate these characteristics in their ability to distinguish between advocates and random users posting on the issue.

4. EVALUATING STRATEGIES

In this section, we describe the datasets used to evaluate our characterizations of advocates of political campaigns in social media. We have two datasets from Twitter, each related to a political campaign carried out using Twitter. We then use these datasets to evaluate the ability of the proposed characterizations of strategies to distinguish between advocates for a given political campaign and random users posting on issues related to the campaign.

4.1 Datasets

We have two datasets related each related to a political campaign. The first dataset is focused on advocates for the Indian election campaign. The rise of around 200 million users of social media in India has made it an important platform for political discourse during elections [20]. Independent groups like NaMo Brigade [8] were formed to advocate for the political campaign of Narendra Modi. The second dataset is related to the issue of gun rights in the United States. This campaign is focused on preserving gun rights, which is being questioned in the wake of increasing gun-related violence. Organizations advocating to preserve gun rights as the National Rifles Association (NRA) have considerable social media presence [30].

Parameter	Elections	Gun Rights
Total # of Users	9390	7695
# of Tweets	20,362,442	19,275,481
# of Links	514,501	899,535
Users posting on the campaign	8500	7000
# of Advocates	890	695

Table 1: Statistics of the datasets of advocates.

Although these organizations have a considerable media presence, it is a challenge to obtain labels for individual users involved in advocacy. Previous literature proposes the use of publicly compiled lists as an effective alternative for inferring affiliations of social media users [21]. The authors in [5] have utilized lists in Twitter to characterize topical-identity based groups. Informed by this literature, we identified two public Twitter lists, titled "NaMo Brigade" [4], for advocates for the election campaign of Narendra Modi and "NRA" [35], for the advocates for gun rights.

To validate the datasets, we apply the mark and recapture technique, drawing from population estimation methodologies [9]. For each of the two lists, we draw two random samples and estimate the total number of errors in the list as follows. Let the probability of finding errors in the random samples r_1 and r_2 be p_{r_1} and p_{r_2}. The number of errors in both the samples will then be given by $e_{r_1} = p_{r_1} N_e$ and $e_{r_2} = p_{r_2} N_e$, where N_e is the total number of errors to be estimated. The number of errors in the intersection of the two samples is then $e_{r_1 r_2} = p_{r_1} p_{r_2} N_e$. The number of errors in the dataset N_e can then be estimated as $N_e = e_{r_1} e_{r_2} / e_{r_1 r_2}$ This is shown to overestimate the population, and hence we use a variation to estimate the number of errors as

$$N_e = \frac{(e_{r_1} + 1) \times (e_{r_2} + 1)}{e_{r_1 r_2} + 1} - 1 \tag{1}$$

A measure of uncertainty is given by the standard error, which estimates of the variability of N if the above experiment is conducted repeatedly. This is computed as follows

$$SE = \sqrt{\frac{(e_{r_1} + 1) \times (e_{r_2} + 1) \times (e_{r_2} - e_{r_1 r_2}) \times (e_{r_1} - e_{r_1 r_2})}{(e_{r_1 r_2} + 1)^2 \times (e_{r_1 r_2} + 2)}} \tag{2}$$

From the standard error, we then calculate the 95% confidence interval i.e. the range within which the number of errors lies with 95% certainty as $I_e = N_e \pm 1.96 \times SE$, where I_e is the 95% confidence interval of the error estimate.

We draw random samples of 10% of the size of the list and use external evaluators to verify them. The evaluators are asked to assess whether a user is an advocate of a given political campaign in social media. The definition of advocates is given to "individuals who use social media to advance strategically their agenda for a given political campaign", according to the definition provided in Section 1. The evaluators mark 1 if they think the member is an advo-

Factors	Features	Elections	Gun Rights
Message Strategy			
Persuasion	Reason	3.52*	6.52**
	Liking	1.717***	1.04***
Focus	Entropy	-3.63***	-0.21**
Emotion	+ve	0.5731**	0.3669**
	-ve	0.7782***	0.2621**
Shared Language		3.8571***	10.57***
Propagation Strategy	Targeting	2.15***	1.91*
	Hubs	2.91***	-.3185
	Authorities	6.39***	0.51*
Community Structure	Following	1.5625***	0.9086***
	Followers	2.045*	0.9554***
	Interactions	1.441***	0.0112**
	Multiple	.468***	4.08***

Table 2: Evaluating strategies using logistic regression coefficients with p-value from t test (*-$p < 0.05$,-$p < 0.01$,***-$p < 0.0001$)**

cate and 0 otherwise. We then estimate the total number of errors from Eq 1 and the confidence interval. The percentage accuracy of the two lists 'NaMo Brigade' and 'NRA' are 92.10%±5.29% and 90.07%±6.27% respectively. Validating the accuracy of the lists, we use their members as ground truths for the set of advocates **a**.

To construct the set **v**, we collect a set of random users who posted with hashtags related to the given campaigns. We obtain the related hashtags from [34] by giving the initial hashtag as "#modi" for the dataset related to the elections and "#progun" for the dataset related to gun rights and assign the set of users posting using the hashtags as **v**. We randomly sampled **v** and a very few number of users were ascertained as advocates, which we removed from the set. We collect the friends, followers, profile, and statuses of both advocates in **a** and random users in **v** for the two campaigns. Some statistics of the datasets is presented in Table 1.

4.2 Evaluation

We now evaluate the characterizations of advocates to distinguish between advocates and random users posting on the issue using the datasets. Characterizations restricted to individual advocates such as persuasion, focus, emotion and propagation strategies are evaluated using logistic regression. The positive class is given by the advocates from the set **a** and the negative class consists of the random users posting on the campaign from **v**. The regression coefficients for all the characterizations along with the significance values derived from the t-test are shown in Table 2.

Characterizations of pairwise relational patterns such as shared language and community structure are evaluated using paired t-test. We present the coefficients of the t-test along with its significance values in Table 2. We next evaluate the characterizations grouping them into message strategies, propagation strategies, and community structure.

4.2.1 Message Strategies

We first evaluate the ability of characterizations of message strategies to distinguish between advocates and random users posting on the campaign. From Table 2, we can see a strong evidence of the use of persuasive language by advo-

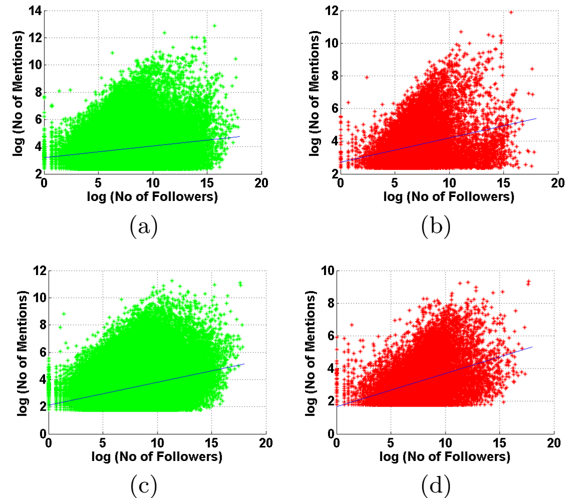


Figure 2: Interacting with Influencers (a) random users posting on the election campaign ($\rho = 0.08$) **(b) advocates for election campaign** ($\rho = 0.15$) **(c) random users posting on gun rights** ($\rho = 0.15$) **(d) advocates for gun rights** ($\rho = 0.20$)

cates in both the datasets, indicative of the attempts of advocates to sway the opinions of others. A significantly higher use of words denoting reason and affinity by advocates can be observed, demonstrating that the proposed model of persuasive language is effective. High positive and negative emotional content, which can be used to generate strong feelings about the campaign, is a strong characteristic in the tweets of advocates. This indicates that unlike workers who are paid to comment or post on a particular issue leading to a lack of emotion in their posts [24], advocates show a higher level of emotion in their messages.

A negative coefficient in the topical focus of advocates with high significance demonstrates that they concentrate their posts around fewer topics for effective messaging compared to random users posting on the campaign. In our experiments, we assign the number of topics t=20. The speech code theory states that users with common affiliations develop a common lingo to foster easier communication [32]. This is borne out by a significantly high coefficient of common hashtags, showing that advocates share higher amount of hashtags with other advocates than with random users posting on the campaign.

These observations indicate that the characterizations of message strategies drawn from theoretical constructs from sociological literature and the proposed approaches to modeling them are effective in distinguishing between advocates and random users posting on the campaign. We next evaluate the ability of the characterizations of the propagation strategies of advocates to distinguish between advocates and random users posting on the campaign.

4.2.2 Propagation Strategies

We first focus on the patterns arising from advocates targeting specific users, which can be performed in Twitter through the mention feature. The number of people who connect to the set of people targeted by the users in **u**, given by **c** in Section 3.2 can be measured by their follower count in Twitter. We first examine if the targeting patterns of advocates differ from those of normal users posting on the issue. We plot the targeting patterns of users in Fig 2 with

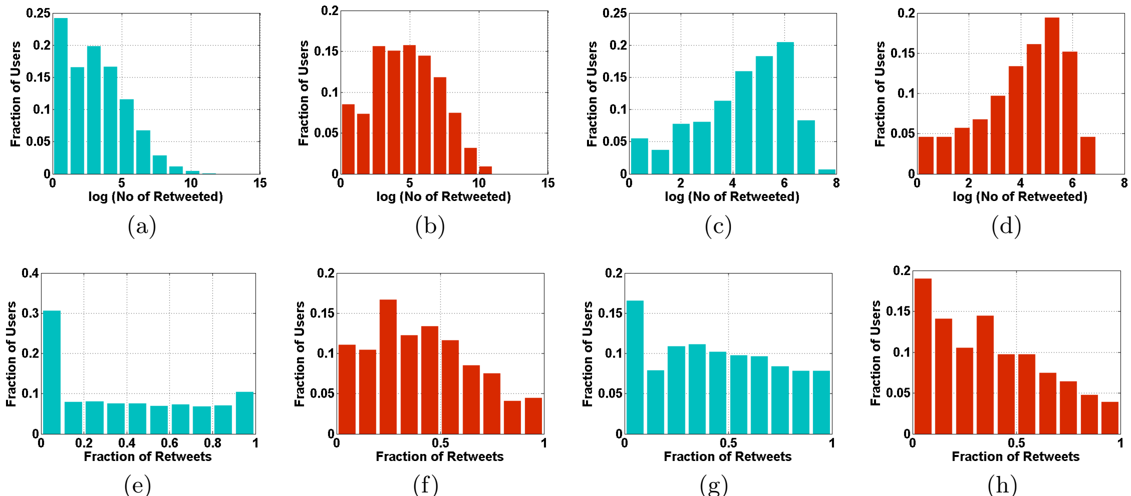


Figure 3: Retweet Patterns. Fraction of Users v/s Fractions of retweets in status for (a) random users posting on and (b) advocates for the election campaign, (c) random users posting on and (d) advocates related for gun rights. Fraction of users v/s no of times users are retweeted by (e) random users posting on and (f) advocates for election campaigns, (g) random users posting on, and (h) advocates for gun rights

the log of the number of mentions on the y-axis and the log of the number of followers on the x-axis. For each plot, we fit a line to the data that minimizes the least square error in the data and its slope, ρ, given in the caption. A higher value of ρ indicates that people with a higher number of followers are targeted more frequently.

We can see that mentioning patterns of advocates from Fig 2 that Fig 2(b) and 2(d) have a greater slope than that of random users posting on the campaign Fig 2(a) and 2(c). Advocates might target popular users as they have the potential to increase the reach of information. We model this notion in Section 3.2 and present the results in Table 2. We can say from a significantly high coefficient value of the targeting characteristic that the targeting patterns of advocates are more significantly skewed towards popular users than those of random users posting on the campaign.

We next evaluate the characterizations of co-propagation behavior of users in $\mathbf{u}$. Fig 3 (a), (b), (c), and (d) illustrates the number of times each user has been retweeted. The logarithm of the times the user is retweeted in shown in the x-axis and the fraction of users who have been retweeted the corresponding number of times, choosing users with more than 100 tweets, on the y-axis. A larger fraction of advocates is retweeted than random users posting on the campaign.

For each user in $\mathbf{u}$, we next compute the ratio of a number of their retweets to his total number of tweets. The retweet ratio is plotted on the x-axis and the fraction of users with the retweet ratio on the y-axis, choosing users with more than 100 tweets, in Fig 3 (e), (f), (g), and (h). A similar pattern can be observed, where the retweet ratio of random users posting on the campaign follows a power law distribution and those of advocates have a skewed distribution with a higher fraction of users actively involved in information propagation. These observations are reflected in Table 2, in the effectiveness of hubs and authority scores for distinguishing advocates and random users posting on the campaign.

These observations indicate that the proposed characterizations of propagation strategies drawn from sociological theories and approaches for modeling them can effectively distinguish between advocates and random users posting on the campaign. We next evaluate the ability of the characterizations of the community structure in their ability to distinguish between advocates for political campaigns and random users posting on the campaign.

4.2.3 Community Structure

We first evaluate the characteristics of social relationships of advocates. From Table 2, we can see that advocates tend to connect with each other significantly more than with random users posting on the topic. This indicates advocates are utilizing social media to build a strong network of connections with each other. Advocates interact significantly more with other advocates for the campaign than with random users posting on it, indicating that they maintain a high level of interactions with each other. These provide evidence of a community structure between advocates.

We next examine if advocates tend to establish multiple types of relationships with each other. From Table 2, we see that given that an advocate has established one type of relation with a user, he has a significantly higher propensity to form multiple types of relation if the user is another advocate than if he is a normal user posting on the campaign. These results indicate a strong network of social connections and interactions between advocates, verifying our postulates. This also provides a basis to infer jointly community membership of users across different relationship types.

Until now, we proposed a set of characteristics capable of effectively distinguishing between advocates and random users posting on the issue. The set of characterizations include both individual characteristics from their messages and propagation strategies and multiple relational patterns like social connections, shared content patterns, and interactions, forming a heterogeneous feature space. We next design a mathematical formulation to combine individual and multiple pairwise characteristics in a unified framework to identify advocates for a political campaign.

5. A UNIFIED MODEL

Integration of individual and relational characteristics in a unified homogenous space for classification can be performed by deriving latent variables from the relational matrix [36].

A difference here is that we have multiple types of pairwise relations between users, and we need to derive latent dimension memberships by jointly exploiting pairwise connections across multiple types of relations. Individual characteristics can be then be combined with the latent variables derived from the relational characteristics to construct features for classification.

Let the proposed individual characteristics be denoted as $\mathcal{I}$ and the relational characteristics be arranged in a tensor $\mathcal{Z}$ as defined in Section 2. Tensors have been used in literature to analyze jointly multi-modal relationships in applications such as community detection [26] and link prediction [12]. We first factorize the tensor $\mathcal{Z}$ to derive latent dimension memberships from the relational characteristics. Different types of connections between two users can be captured by their similarity in their memberships of latent dimensions. Two users who have pairwise relationships with each other across different types will have higher similarity in their latent dimension memberships than two users who do not have pairwise relationships with each other.

We factorize the tensor $\mathcal{Z}$ to obtain the user matrices $\mathbf{U} \in \mathbb{R}^{\mathrm{N}\times\mathrm{K}}$ and $\mathbf{V} \in \mathbb{R}^{\mathrm{N}\times\mathrm{K}}$, and the relationship type dimension matrix $\mathbf{T} \in \mathbb{R}^{\mathrm{T}\times\mathrm{K}}$, where K is the number of latent dimensions, by solving the following optimization problem

$$\min_{\mathbf{U},\mathbf{V},\mathbf{T}} ||\mathcal{Z} - [\![\mathbf{U},\mathbf{V},\mathbf{T}]\!]||_{\mathrm{F}}^2, \tag{3}$$

where $[\![\mathbf{U},\mathbf{V},\mathbf{T}]\!] \in \mathbb{R}^{\mathrm{N}\times\mathrm{N}\times\mathrm{T}}$ is given by

$$[\![\mathbf{U},\mathbf{V},\mathbf{T}]\!] = \sum_{k=1}^{K} \mathbf{u}_k \circ \mathbf{v}_k \circ \mathbf{t}_k.$$

Here $\mathbf{u}_k$, $\mathbf{v}_k$, $\mathbf{t}_k$ are the k^{th} column vectors of $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{T}$ respectively. The symbol $\circ$ represents the vector outer product such that if the tensor $\mathcal{Y} = \mathbf{u}_k \circ \mathbf{v}_k \circ \mathbf{t}_k$ then $\mathcal{Y}_{\mathrm{efg}} = (\mathbf{u}_k)_{\mathrm{e}}(\mathbf{v}_k)_{\mathrm{f}}(\mathbf{t}_k)_{\mathrm{g}}$. Substituting this in Eqn 3, we get

$$f = \min_{\mathbf{U},\mathbf{V},\mathbf{T}} ||\mathcal{Z} - \sum_{k=1}^{K} \mathbf{u}_k \circ \mathbf{v}_k \circ \mathbf{t}_k||_{\mathrm{F}}^2, \tag{4}$$

We optimize this function motivated by the conjugate linear optimization method [1]. We first arrange the vectors $\mathbf{u}$,$\mathbf{v}$ and $\mathbf{t}$ in a single vector $\mathbf{x} = [\mathbf{u},\mathbf{v},\mathbf{t}]$ and calculate the gradient of $\mathbf{f}(\mathbf{x})$ with respect to each $\mathbf{x}_k^n$ where $\mathbf{x}^1 = \mathbf{u}, \mathbf{x}^2 = \mathbf{v}, \mathbf{x}^3 = \mathbf{t}$. f can be rewritten as

$$f = \underbrace{\frac{1}{2}||\mathcal{Z}||^2}_{f_1} - \underbrace{\langle \mathcal{Z}, \sum_{k=1}^{K} \mathbf{x}_k^1 \circ \mathbf{x}_k^2 \circ \mathbf{x}_k^3 \rangle}_{f_2} + \underbrace{\frac{1}{2}|| \sum_{k=1}^{K} \mathbf{x}_k^1 \circ \mathbf{x}_k^2 \circ \mathbf{x}_k^3 ||^2}_{f_3} \tag{5}$$

The gradient of f is obtained by computing its partial derivative with respect to each element in $\mathbf{x}$ denoted by $\mathbf{x}_k^m$. So

$$\frac{\partial f_1}{\partial \mathbf{x}_k^m} = 0 \tag{6}$$

as $\mathcal{Z}$ is a constant with respect to $\mathbf{x}_k^m$. The partial derivative of f_2 with respect to each element in $\mathbf{x}$ denoted by $\mathbf{x}_k^m$ can then be computed as follows.

$$\frac{\partial f_2}{\partial \mathbf{x}_k^1} = \mathcal{Z} \times_2 \mathbf{x}_k^2 \times_3 \mathbf{x}_k^3, \frac{\partial f_2}{\partial \mathbf{x}_k^2} = \mathcal{Z} \times_1 \mathbf{x}_k^1 \times_3 \mathbf{x}_k^3, \frac{\partial f_2}{\partial \mathbf{x}_k^3} = \mathcal{Z} \times_1 \mathbf{x}_k^1 \times_2 \mathbf{x}_k^2, \tag{7}$$

where $\times_n$ is the n-mode multiplication operator as defined as Section 2. The partial derivative of f_3 with respect to each term in $\mathbf{x}$ denoted by $\mathbf{x}_k^m$ can be computed as

Method	Elections		Gun Rights	
	AUC	**F1**	**AUC**	**F1**
Random	0.4983	0.1607	0.5053	0.1536
Retweet Ratio	0.5804	0.1797	0.5078	0.1118
Volume	0.6830	0.2406	0.6519	0.2332
Bag of Words	0.7379	0.3515	0.7305	0.2919
Combine	0.7599	0.3604	0.7460	0.3065
Our Method	**0.9301**	**0.6341**	**0.9431**	**0.6046**

Table 3: Comparison with different baselines.

$$\frac{\partial f_3}{\partial \mathbf{x}_k^m} = \sum_{j=1}^{K} (\prod_{\substack{r=1 \\ r\neq m}} \mathbf{x}_k^{r\mathrm{T}} \mathbf{x}_j^r) \mathbf{x}_j^r \tag{8}$$

The overall gradient can then be computed as

$$\frac{\partial f}{\partial \mathbf{x}_k^m} = \frac{\partial f_1}{\partial \mathbf{x}_k^m} - \frac{\partial f_2}{\partial \mathbf{x}_k^m} + \frac{\partial f_3}{\partial \mathbf{x}_k^m} \tag{9}$$

where $\frac{\partial f_1}{\partial \mathbf{x}_k^m}$, $\frac{\partial f_2}{\partial \mathbf{x}_k^m}$, $\frac{\partial f_3}{\partial \mathbf{x}_k^m}$ are as described in Eqn 6, Eqn 7 and Eqn 8 respectively. The gradient descent step repeated for all values of m and k is continued until convergence. As the objective function in Eqn 4 is convex, the optimization is guaranteed to converge. The computational complexity of the iterations is low due to the high sparsity of $\mathcal{Z}$.

To give an intuitive understanding of the optimization term in Eqn 4, we rewrite it as follows

$$\min_{\mathbf{U},\mathbf{V},\mathbf{T}} || \sum_{\mathrm{t}=0}^{\mathrm{T}} \mathcal{Z}_{\mathrm{t}} - \mathbf{U}\mathbf{D}_{\mathrm{t}}\mathbf{V}^{\mathbf{T}} ||_F^2, \tag{10}$$

where $\mathcal{Z}_{\mathrm{t}}$ represents the user relations of type t and $\mathbf{D}_{\mathrm{t}} \in \mathbb{R}^{\mathrm{K}\times\mathrm{K}}$ is a diagonal matrix whose diagonal elements are the t^{th} row of $\mathbf{T}$. The matrices $\mathbf{U}$ and $\mathbf{V}$ contains the latent dimension memberships of users jointly inferred across different relationship types. If $\mathcal{Z}_t \forall t$ is symmetric then $\mathbf{U} = \mathbf{V}$. The matrix $\mathbf{T}$ contains the contribution of each relation type to different dimensions. For example, a high value of $\mathbf{t_2}$ signifies that high-scoring users in $\mathbf{u_2}$ form connections with high-scoring users in $\mathbf{v_2}$ through relationship type 2. The latent features representing different kinds of pairwise relationships between users can be obtained from any linear combination of $\mathbf{U}$ and $\mathbf{V}$. For our experiments, we use $\mathbf{L} = \mathbf{U} + \mathbf{V}$ after column normalization as latent features. We combine the individual characteristics $\mathcal{I}$ and the latent features $\mathbf{L}$ to construct a feature set $F = \{\mathcal{I}, \mathbf{L}\}$ for identifying advocates for political campaigns in social media.

6. IDENTIFYING ADVOCATES

In this section, we evaluate the performance of our framework by answering the following questions. How effective is our framework for identifying advocates for political campaigns on social media? How good are the characteristics group in identifying advocates? How robust is the proposed framework for variation in training sizes?

6.1 Performance Evaluation

We classify the feature set derived in Section 5 using Linear Discriminant Analysis and perform 10-fold cross validation to evaluate the performance of the framework in identifying advocates. We measure the performance using two metrics, AUC, and F1-measure and present the results in

Table 3. We have the following baselines to compare the performance of our framework.

- **Random :** We randomly assign labels to all the users.
- **Retweet Ratio [27]:** The fraction of retweets per overall tweets is used as a feature.
- **Activity :** The total number of tweets of the users, is used for as a feature.
- **Bag of Words :** We use all the words in the status messages of users as features after tf-idf weighting,
- **Combine:** We combine all the proposed baselines by concatenating all the features.

We compare the performance of the baselines and the proposed framework and illustrate the results in Table 3. Random assignment gives an AUC value of around 0.5 and the F1 measure is low for both the datasets indicating the difficulty of the problem. The retweet ratio is used to characterize the propagating behavior of biased users in [27]. We model a wider range of propagation strategies advocates employ in social media, and as we can see from Table 4, where our characterizations of propagation strategies outperform the retweet ratio. From Table 3, we see that advocates are more active than random users posting on the campaign. We model specific patterns in the messages based on the strategies of advocates and hence outperform this baseline that considers only the total number of messages.

The "Bag of Words" performs better than the other baselines, but the number of features here is high. We model specific strategies of advocates related to the message, propagation and community structure instead of using all the posted words, enabling us to outperform this baseline. Combining all the baselines gives a slight improvement in the performance, indicating the potential benefits of integrating heterogeneous information. The proposed framework outperforms the baselines demonstrating that it effectively models and integrates characteristics useful for identifying advocates for political campaigns. This signifies the ability of the framework in understanding the strategies and model them to effectively to identify advocates. We perform the t-test between the results of our framework and the baselines and find that the difference is significant.

In summary, we can say that our framework outperforms the baselines demonstrating that it effectively models and integrates strategies useful for identifying advocates for political campaigns. We next analyze the contributions of different characteristic groups in identifying advocates.

6.2 Contributions of Characteristic Groups

We separately select characteristics related to message strategies, propagation strategies, and community structure and present them to the classifier. We compare the performance of different characteristic groups in identifying advocates using AUC and F1 measure with 10-fold cross validation and illustrate the results in Table 4.

We first examine the performance of message strategies by combining individual characteristics and the latent features from shared content. From Table 4, we can see that the characterizations of message strategies like persuasion, emotion, focus, and shared linguistic patterns outperform random characteristics by a significant margin. This demonstrates that proposed characterizations of messages strategies contribute significantly in identifying advocates.

Method	**Elections**		**Gun Rights**	
	AUC	**F1**	**AUC**	**F1**
Random	0.4983	0.1607	0.5053	0.1536
Mess Strat	0.8240	0.4303	0.8517	0.4727
Prop Strat	0.8210	0.3689	0.5707	0.1904
Mess+ Prop Strat	0.8804	0.5152	0.8680	0.4934
Comm Struct	0.8918	0.5117	0.8859	0.4834
Overall	0.9301	0.6341	0.9431	0.6046

Table 4: Performance of different groups

The performance of characteristics related to propagation strategies is much higher in the dataset related to elections than in the dataset related to gun rights. This is an indication that advocates in election campaigns place more emphasis on information propagation. On combining characterizations from both message strategies with propagation strategies, we observe an improvement in performance. This demonstrates the contribution of characterizations of propagation strategies in identifying advocates.

The characteristics related to community structure perform well in both the datasets. This indicates that advocates in social media have strong relationships with other advocates of the issue and display strong interactions with each other. A combination of all the three characteristic groups performs significantly better than individual characteristic groups demonstrating the effectiveness of our framework for integrating these characterizations.

In summary, we can say that all the proposed characteristic groups contribute significantly to identifying advocates for political campaigns in social media. We next evaluate the robustness of the framework to variations of training data size and assess its effect on identifying advocates.

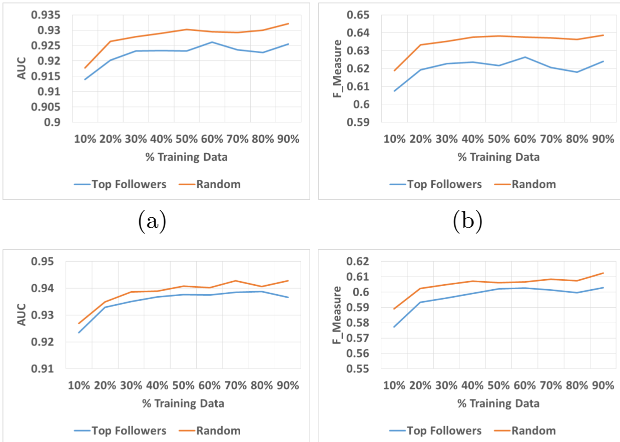
6.3 Performance with Varying Training Sizes

We now answer the following questions: How does the framework perform with varying proportions of training data? How effective is the framework when we use the labels of popular users, potentially more accessible, for training?

We vary the relative proportion of training and testing sizes from $\alpha = \{10\% - 90\%\}$ with increments of 10%. For each value of α, we take the mean performance of 100 random samples with $\alpha\%$ used for training and the rest for testing. We repeat the procedure for the different values of α and illustrate the results in Figure 4.

In many cases, the labels of popular users might be well known, and the labels might be therefore potentially easier to obtain. To evaluate the framework in this scenario, we take the number of followers of a user as an indication of his popularity and sort the users in decreasing number of their followers. For each value of α, we select the users in the top $\alpha\%$ of followers for training and the rest for testing and obtain the performance. We repeat the procedure for all the values of α and illustrate the results in Figure 4.We make the following observations from the figure.

The performance of the algorithm is significantly higher than the nearest baseline for all relative proportions of training data for both methods of sampling. The performance slightly increases for higher proportions of training data but is overall robust for varying size of training data. Finally, random selection of training data performs only slightly better than selecting users with a high number of followers. This demonstrates that labels of just the popular users, which are



(a) (b) (c) (d)

Figure 4: Effect of size of training data with training data randomly chosen and chosen ordered according to the number of followers for election dataset with (a) AUC (b) F1 measure and dataset on the gun rights with (c) AUC and (d) F1 Measure

potentially easier to obtain, can be effective in identifying advocates for political campaigns on social media.

In summary, the experiment demonstrates that the framework is robust to variation in training data size and can also effectively identify advocates for political campaigns even when only the labels of popular users are known.

7. RELATED WORK

We present the related work in three categories: the study of advocacy groups in social media, the use of social media to carry out campaigns and identification of attributes of social media users. We next survey each of them in detail.

Advocacy in social media has received considerable attention in the recent sociological literature. Obar et.al [29] surveyed advocacy organizations to analyze their social media use and stated that the ability of social media to facilitate civic engagement and foster collective action is the primary motivations. The Twitter use of nonprofit organizations was studied in [17]; communicating relevant information, building user communities and call for action were found to be their principal objectives. The behavior of advocacy groups during crises was investigated in [38], and the role of social media in organizing and distributing work related to rescue operations was discussed. The Facebook pages of advocacy groups were studied in [7], and they link behaviors such as network activity and query responses by advocacy groups to the extent and growth of their network. These works provide a qualitative understanding of the activities of advocacy organizations in social media, but do not propose models to identify individual advocates.

The use of social media to carry out campaigns has received considerable attention in literature [18, 23]. Extracting campaigns from social media using textual similarity has been studied in [23] and the authors here do not concentrate on identifying individual advocates. Crowdturfing, where paid crowdsourced workers are used to post advertisements for a product, has generated considerable interest [24, 10, 25]. Paid workers are detected by less use of emotion in their content [24], predominant use of URLs [25] and lack of interaction with other users [10]. We find that advocates in social media have high emotional and textual content in their messages and show strong communication among themselves. Substantial work has been done on the behavior of opinion leaders [3] in social media. Any social media user can try to advance an agenda on an issue without necessarily being an opinion leader.

Researchers have studied the spread of political misinformation [33] by identifying malicious links that spread through social networks. We are interested in individual users who are advocating for an issue, and they need not necessarily spread misinformation. The behavior of social media users with bias on a given political issue have been studied in [27]. The paper identifies users who consistently retweet messages from users with a particular viewpoint. We aim to identify users who actively try to advance an agenda that manifests in their distinct strategies.

Substantial efforts have been made to identify attributes of social media users to give an additional context to their posts. The properties of social relationships is shown to be beneficial in identifying user attributes across social media sites [40]. Researchers have worked on identifying political orientations of users from their retweets [39] and network connections [2]. The architecture of a web-based system to aggregate profile attributes of a user drawn from diverse social media platforms is described in [16]. These works focus on inferring different personality attributes of social media users but do not determine whether a user is an advocate for a political campaign on social media.

8. CONCLUSIONS

In this paper, we present a framework to identify advocates for political campaigns on social media. We characterize advocates through their message strategies, propagation strategies, and community structure and propose different characterizations based on them. We integrate heterogeneous information derived from these diverse characterizations and demonstrate that the framework can identify effectively advocates for political campaigns on social media. We analyze contributions of individual characteristic groups like message strategies, propagation strategies, and community structure in identifying advocates. Finally, we evaluate the performance of the framework for different proportions of training sizes and demonstrate that the framework is robust to variation in training data size. This research is in a step towards informing users before they imbibe information from these accounts and also enabling campaign managers to identify advocates to assist in digital political campaigns.

Future work can consider the detection of groups of people involved in advocacy in social media drawing from community detection methods. Advocates might use strategies other than those studied here, and investigation into the strategies will increase the understanding of their behavior. The impact of advocates on meme popularity in social media is an interesting area of future research and can give insight on the impact of their activities. Studying information seeking patterns among social media advocates of a particular political campaign can give interesting insights into interactions among members of social communities. Finally, studying the evolution of the behavior and interaction networks of advocates can provide information about the process of community formation in online social media.

9. ACKNOWLEDGEMENTS

This material is based upon work supported by, or in part by, Office of Naval Research (ONR) under grant number N000141010091.

10. REFERENCES

[1] E. Acar, D. M. Dunlavy, and T. G. Kolda. A scalable optimization approach for fitting canonical tensor decompositions. *J.Chemo*, 2011.

[2] L. Akoglu. Quantifying political polarity based on bipartite opinion networks. In *ICWSM*, 2014.

[3] S. Aral and D. Walker. Identifying influential and susceptible members of social networks. *Science*, 337(6092):337–341, 2012.

[4] A. Baruah. Namo Brigade. http://goo.gl/JKdmlP, 2014. [Online; accessed 7-May-2015].

[5] P. Bhattacharya, S. Ghosh, J. Kulshrestha, M. Mondal, M. B. Zafar, N. Ganguly, and K. P. Gummadi. Deep twitter diving: exploring topical groups in microblogs at scale. In *CSCW*. ACM, 2014.

[6] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *JMLR*, 2003.

[7] D. S. Bortree and T. Seltzer. Dialogic strategies and outcomes: An analysis of environmental advocacy groups' facebook profiles. *PRR*, 2009.

[8] N. Brigade. Namo Brigade, Mission-Narendra Modi for PM. http://goo.gl/dNbKhd, 2014. [Online; accessed 9-November-2014].

[9] J. E. Brower, J. H. Zar, and C. N. Von Ende. Field and laboratory methods for general ecology. 1998.

[10] C. Chen, K. Wu, V. Srinivasan, and X. Zhang. Battling the internet water army: Detection of hidden paid posters. In *ASONAM*, 2013.

[11] R. B. Cialdini. *Influence: The psychology of persuasion*. Quill New York, NY, 1993.

[12] D. M. Dunlavy, T. G. Kolda, and E. Acar. Temporal link prediction using matrix and tensor factorizations. *TKDD*, 2011.

[13] D. M. Farrell and P. Webb. Political parties as campaign organizations. 2006.

[14] I. Gilbert and T. Henry. Persuasion detection in conversation. Technical report, DTIC Document, 2010.

[15] S. González-Bailón, J. Borge-Holthoefer, and Y. Moreno. Broadcasters and hidden influentials in online protest diffusion. *American Behavioral Scientist*, page 0002764213479371, 2013.

[16] P. Gundecha, S. Ranganath, Z. Feng, and H. Liu. A tool for collecting provenance data in social media. In *KDD*, pages 1462–1465. ACM, 2013.

[17] C. Guo and G. D. Saxton. Tweeting social change: How social media are changing nonprofit advocacy. *NVS*, 2013.

[18] X. Hu, J. Tang, Y. Zhang, and H. Liu. Social spammer detection in microblogging. In *IJCAI*, 2013.

[19] D. H. Jernigan and P. A. Wright. Media advocacy: Lessons from community experiences. *JPHP*, 1996.

[20] A. Khullar and A. Haridasan. Politicians slug it out in India's first social media election. http://goo.gl/ZXZm3W, 2014. [Online; accessed 9-November-2014].

[21] D. Kim, Y. Jo, I.-C. Moon, and A. Oh. Analysis of twitter lists as a potential source for discovering latent characteristics of users. In *CHI Workshop on Microblogging*, 2010.

[22] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *JACM*, 1999.

[23] K. Lee, J. Caverlee, Z. Cheng, and D. Z. Sui. Campaign extraction from social media. *TIST*, 2014.

[24] K. Lee, P. Tamilarasan, and J. Caverlee. Crowdturfers, campaigns, and social media: Tracking and revealing crowdsourced manipulation of social media. In *ICWSM*, 2013.

[25] H. Li, A. Mukherjee, B. Liu, R. Kornfield, and S. Emery. Detecting campaign promoters on twitter using markov random fields. In *ICDM 2014*, pages 290–299, Dec 2014.

[26] Y.-R. Lin, J. Sun, H. Sundaram, A. Kelliher, P. Castro, and R. Konuru. Community discovery via metagraph factorization. *TKDD*, 2011.

[27] C. Lumezanu, N. Feamster, and H. Klein. # bias: Measuring the tweeting behavior of propagandists. In *ICWSM*, 2012.

[28] J. D. McCarthy and M. N. Zald. Resource mobilization and social movements: A partial theory. *AJS*, 1977.

[29] J. A. Obar, P. Zube, and C. Lampe. 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. *JIP*, 2012.

[30] A. Palmer. NRA campaign in high gear. http://goo.gl/PYBXDq, 2014. [Online; accessed 9-November-2014].

[31] J. W. Pennebaker, M. E. Francis, and R. J. Booth. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 2001.

[32] G. Philipsen, G. Philipsen, and T. Albrecht. A theory of speech codes. *Dev.CommTheo*, 6, 1997.

[33] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, A. Flammini, and F. Menczer. Detecting and tracking political abuse in social media. In *ICWSM*, 2011.

[34] RiteTag. Twitter hashtags. http://goo.gl/X88aY7, 2014. [Online; accessed 7-August-2014].

[35] M. Robinson. NRA, PRO-GUN, GUN OWNER. http://goo.gl/OyDDGR, 2014. [Online; accessed 7-August-2014].

[36] L. Tang and H. Liu. Relational learning via latent social dimensions. In *KDD*, 2009.

[37] I. Today. Rise of the Cyber Hindu. http://goo.gl/NvxAw6, 2014. [Online; accessed 7-May-2015].

[38] J. I. White, L. Palen, and K. M. Anderson. Digital mobilization in disaster response: Self-organization of on-line pet advocates in response to hurricane sandy. In *CSCW*, 2014.

[39] F. M. F. Wong, C. W. Tan, S. Sen, and M. Chiang. Quantifying political leaning from tweets and retweets. In *ICWSM*, 2013.

[40] R. Zafarani and H. Liu. Users joining multiple sites: Friendship and popularity variations across sites. *Information Fusion*, 28:83 – 89, 2016.