

Social Network of Extreme Tweeters: A Case Study

Xiuwen Zheng
 San Diego Supercomputer Center
 University of California San Diego
 La Jolla, CA 92093
 xiz675@eng.ucsd.edu

Amarnath Gupta
 San Diego Supercomputer Center
 University of California San Diego
 La Jolla, CA 92093
 a1gupta@ucsd.edu

Abstract—The number of posts made by a single user account on a social media platform Twitter in any given time interval is usually low. However, there is a subset of users whose volume of posts is much higher than the median. In this paper, we investigate the content diversity and the social neighborhood of these extreme users and others. We define a metric called “interest narrowness”, and identify that a subset of extreme users, termed anomalous users, post with very low topic diversity. We show that anomalous groups have the strongest within-group interactions, compared to their interaction with others, and exhibit different information sharing behaviors with other anomalous users compared to non-anomalous extreme tweeters.

Index Terms—Twitter, Social Media, user characterization, network analysis, content diversity, behavior analysis

I. INTRODUCTION

Social Media is part of our daily lives, and increasingly more people are actively participating in various social media platforms. It was recently reported [1] that Twitter has 126 million daily users, with an estimated annual growth rate of about 9%. It is estimated that roughly 46% of Twitter users are on the platform daily. In this paper, we investigate the following questions: *What types of users tweet an enormous amount and what do they talk about?*

The intuition behind this paper comes from the observation that we can characterize users’ tweeting behavior based on the volume and the content diversity of their tweets.

We first consider tweet volumes for individual users. Based on our data set of over 1.5 billion tweets, we observe that over any arbitrary time interval, the number of tweets by a user follows a power law type of distribution (see Fig. 1(a)) – most users post very few tweets while only a few users write considerably more tweets. In this empirical result from July, 2017, only 20% of the users posted more than 24 tweets. We use the term *extreme tweeters* (ETT) for users who tweet more frequently than an average user in any given time interval.

A different stratification of users can be created based on the *content diversity* (c-diversity) of their posts. Intuitively, a user

who is interested in many topics will have a higher content diversity than a user with a narrow range of interests (e.g., only football). One simple, but rough measure of c-diversity is the number of distinct words (not counting mentions) used by a user over all their posts in a given period of time. Fig. 1(b) shows a typical plot of the c-diversity of users as a function of the number of tweets they sent. In general, users who tweet more tend to have higher c-diversity. Clustering the frequency distribution reveals three different clusters corresponding to (a) users who tweet less and use fewer distinct words (blue cluster), (b) users who tweet more and have higher c-diversity (greenish yellow cluster), and (c) the small number of users who tweet more and yet have low c-diversity (red cluster).

In this case study, we explore the tweeting behavior and the social network of ETT, who constitute groups (b) and (c) above. We are particularly interested in group (c) because at first glance, their high-tweet-rate, low-diversity behavior is anomalous and somewhat counterintuitive. In the light of this exploration, the paper makes the following contributions:

- 1) We propose a novel way to classify users based on their tweet rate and a new measure of verbosity called *interest narrowness*.
- 2) We present an algorithm to detect the anomalous users.
- 3) We investigate the nature of the network relations of the anomalous group and other users.
- 4) We show that the social interaction of anomalous and non-anomalous groups vary with time, and the vigorousness of the interaction intensifies around events like elections.

Note that this paper is a short version of our work and for details please refer to [2].

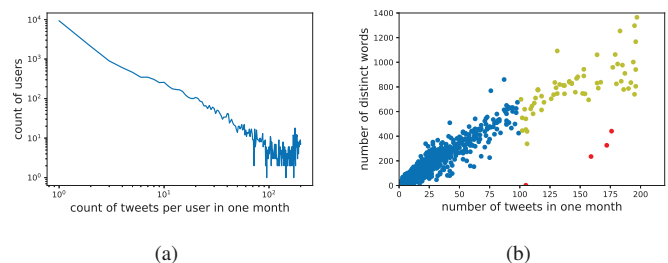


Fig. 1: (a) Frequency distribution of tweet count for users in a month. (b) Frequency distribution of distinct words vs. number of tweets in a month.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ASONAM '19, August 27-30, 2019, Vancouver, Canada

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6868-1/19/08/\$15.00

<https://doi.org/10.1145/3341161.3342909>

II. RELATED WORK

The problem of user characterization in a social network has been investigated by many research groups. We present a few samples of these research efforts.

A 2016 survey [3] covers a wide bandwidth of different approaches for user characterization. The behavioral properties they report range from conscientiousness and extroversion to privacy behavior, deceptive traits and response to social attacks. In contrast to our approach, many of the reported analyses studied in this paper are based on user surveys.

Gabrowicz et al [4] take a bond-theory based approach and distinguish between social user groups and topical user groups in a network based on features like reciprocity, topicality and activity. Of these, topicality, defined based on a metric called “normalized entropy” measures how much the topics of discussion vary within a group. The higher the entropy, the greater is the variety of terms and, according to the theory, the more social the group is. However, this measure considers the words to be independent, which usually does not hold.

Similar to our notion of anomalous users, [5] investigates the concept of “dedicators”, users who transmit information in selected topic areas to the people in their egocentric networks. The concept of dedication is determined by volume, engagement, personal tendencies and topic weight, where personal tendency includes the user’s topic diversity as measured by Latent Dirichlet Allocation (LDA), and engagement measures the activity level of conversations.

Diversity of topics within text posts is also analyzed in papers like [6] that perform LDA to compute topics and then determine topic diversity across a users posts as the number of distinct probable topics found across all of the users posts.

Closer to our application, Bail et al [7] used a POS-tagged BOW model to analyze Facebook posts for political content and derived a network of correlated concepts. They demonstrated how some advocacy organizations produce social media messages that inspire far-ranging conversation among social media users. Interestingly, their network analysis is based on the co-occurrence of concept terms from which they extract interesting connection patterns that characterize influence modalities for advocacy groups.

In contrast to all related work, ours is the first attempt, to our knowledge, that specifically analyzes the posting behavior and interaction patterns of extreme tweeters. Although we use Twitter as our example social media platform, our method is equally valid for any other platform that exhibits high-volume postings and vigorous interactions.

III. THE SETTING

We now formalize the intuitive user behavior classification scheme presented in Section I.

A. Classification based on Tweet Volume

Assume that Δ is the minimal analytic interval (MAI), i.e., a minimum time-interval (e.g., 1 week) over which data is collected in order to perform any user behavior analysis, we define ETT behavior as,

Definition 1 (ETT Behavior): Let U be the set of users who tweet during Δ , and T^u be the number of posts by a user $u \in U$ during Δ , then we say an user u' **exhibits ETT behavior in Δ** if $T^{u'} \geq \mathbf{E}[T^u] + \delta \cdot \sqrt{\mathbf{Var}[T^u]}$, where $\delta \geq 0$ is an arbitrary constant to control the selectivity.

Users with ETT behavior in Δ are called *ETT users* in the MAI, while others are *regular users*.

B. Classification based on Content Diversity

Instead of counting the number of distinct words for a user, which cannot capture the thematic diversity, we develop a c-diversity measure called *Interest Narrowness* by using the bag-of-words and singular value decomposition (SVD) techniques.

For a user u , we construct its text matrix $\mathbf{M}^u$ by adopting the bag-of-words model on all tweets of u during a certain MAI Δ . Let p be the tweet count and q be the number of distinct words (except stop words and URLs) over all tweets of u in this MAI. Apparently, $\mathbf{M}^u$ is of dimension $p \times q$, and without loss of generality, we assume that $p \leq q$. By applying SVD, we have $\mathbf{M}^u = \mathbf{U}\Sigma^u\mathbf{V}$ where the diagonal entries of Σ^u , $\sigma_i^u = \Sigma_{i,i}^u$, are singular values and there are $\min(p, q) = p$ singular values. Equivalently, $\mathbf{M}^u$ can be rewritten by a weighted sum of p separable matrices: $\mathbf{M}^u = \sum_i \mathbf{A}_i^u = \sum_i \sigma_i^u \cdot \mathbf{U}_i \otimes \mathbf{V}_i$, where $\mathbf{U}_i$ and $\mathbf{V}_i$ are the i^{th} columns of $\mathbf{U}$ and $\mathbf{V}$ respectively, and $\otimes$ refers to the outer product. Based on the above decomposition, we define the contribution of j^{th} separable matrix $\mathbf{A}_j^u$ as follows,

$$c_j^u = \frac{\sigma_j^{u2}}{\sum_i \sigma_i^{u2}}. \quad (1)$$

To speed up the computation, we can approximate matrix decomposition by using the randomized SVD [8] where only partial singular values are computed, $\hat{\mathbf{M}}^u \approx \hat{\mathbf{U}}\hat{\Sigma}^u\hat{\mathbf{V}}$, where $\hat{\mathbf{U}}$ is of $p \times k$, $\hat{\mathbf{V}}$ is of $k \times q$ and $\hat{\Sigma}^u$ is of $k \times k$ where $k < \min(p, q)$. The interest narrowness is then given by,

$$\eta^u = \frac{\sum_{j=1}^k \tilde{\sigma}_j^{u2}}{\sum_i \sigma_i^{u2}} = \frac{\sum_{j=1}^k \tilde{\sigma}_j^{u2}}{\|\mathbf{M}^u\|_F^2}, \quad (2)$$

where $\|\mathbf{M}^u\|_F$ is the Frobenius norm of matrix $\mathbf{M}^u$ and $\|\mathbf{M}^u\|_F^2 = \sum_i \sigma_i^{u2}$. The commonly used implementation of randomized SVD takes time $O(pq \log k + (p + q)k^2)$ [9]. Notice that k serves as a hyper-parameter in the computation of narrowness. In our experiments we observe that setting k to $\max(10, p/10)$ can well represent the c-diversity of all tweets. For this fixed k , a larger η stands for narrower topic interests.

C. Framework

Given any MAI Δ , we can define *anomalous users* (shown as the red cluster in Fig. 1(b)) as users with ETT behavior and narrow topic interests during Δ . Algorithm 1 provides the framework to detect anomalous users based on the definition. The algorithm has two hyper-parameters, δ and λ , which jointly control the selectivity of anomalous users. Intuitively, the larger δ and λ are, the more strict the criteria of anomalous users will be. Given a time period, lines 1-3 find all ETT

users in Δ based on Definition 1. Lines 4-6 calculate, for each ETT user, the corresponding interest narrowness. Finally, line 7 finds out users from ETT with high interest narrowness values as anomalous users.

Algorithm 1: Anomalous User Detection

Input: a list of triples: $\mathcal{T} = \{(T.u, T.text, T.time)\}$, a fixed time period: Δ , and hyper-parameters: δ and λ
Output: a list of anomalous users: $\mathcal{A}$

```

1  $\mathcal{T}' \leftarrow \{(T.u, T.text) | T \in \mathcal{T}, T.time \in \Delta\}$ ;
2 group by users to get per user tweet count and tweet corpus as
   $\mathcal{C} \leftarrow \{(C.u, C.count, C.all\_texts)\}$ ;
3  $ETT \leftarrow \{C.u | C \in \mathcal{C}, C.count \geq \mathbf{E}[C.count] + \delta\sqrt{\mathbf{Var}[C.count]}\}$ ;
4  $\mathcal{H} \leftarrow \{\}$ ;
5 for  $u \in ETT$  do
6   use RM (Eq. (2)) to calculate  $nrrw; \mathcal{H}.add((u, nrrw))$ ;
7  $\mathcal{A} \leftarrow \{H.u | H \in \mathcal{H}, H.nrrw \geq \mathbf{E}[\mathcal{H}.nrrw] + \lambda\sqrt{\mathbf{Var}[\mathcal{H}.nrrw]}\}$ ;
```

IV. SOCIAL NETWORK ANALYSIS

A. Rationale

So far, we have focused on the ETT behavior of users and identified narrow-interest users with ETT behavior. We now explore the social network around these users to study their interaction patterns with other users. We would like to investigate four questions about anomalous users:

- (Q1): Do anomalous users interact heavily with each other?
- (Q2): Do anomalous users interact heavily with extreme tweeters or regular users?
- (Q3): If some anomalous users interact heavily with each other, what is their behavior pattern as a group?
- (Q4): Do anomalous users have different behaviors during different time periods, especially around major public-opinion-inciting events?

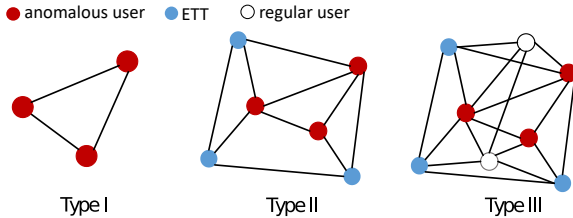


Fig. 2: Three connection patterns around anomalous users.

B. Three Simple Patterns

In Fig. 2 we show three simple connection patterns involving the three user categories. The edges in the graph represent the *mentions* relationship between users and they are undirected because in this case study we are simply exploring the connections and not the direction of messages that flow between user groups.

The Type-I pattern, denoted as $G_1 = (V_1, E_1)$, shows only within-group triads for anomalous users. The edges represent the relationship between two anomalous users during a certain time period T . The Type-II pattern, denoted as $G_2 = (V_2, E_2)$, extends the triads to include first neighbors of anomalous users who are extreme tweeters in T . E_2 also includes the mentions between these non-anomalous ETT users. As for the third case, the Type-III pattern network is formed by considering

anomalous users and *all* their first neighbors. These three patterns correspond to the questions Q1 and Q2.

Let G^T be the complete social network within time interval T . We extract the largest subgraphs from G^T satisfying each connection pattern respectively, called Type- x network ($x = 1, 2, 3$), denoted as G_x^T .

C. Metrics

In this subsection, we propose three measures to assess the connectivity and density of our connection patterns in the network during any time interval. With these measures we expect to answer all the four questions provided.

1) *Coreness Distribution*: k -core decomposition is a standard technique for tasks like dense region detection [10], [11] and network feature extraction [12], [13]. We introduce the concept of “coreness distribution”, which is the distribution of nodes over coreness. To capture the holistic connective characteristic of anomalous users and their related users, we will examine the coreness distributions in three types of social networks $G_x^T, x = 1, 2, 3$.

2) *Measure of Collective Behavior*: The coreness distribution depicts the holistic interaction intensity of all anomalous users with different categories of users, which answers Q1 and Q2. However, it is possible that an anomalous user may not strongly connect with other anomalous users. We focus on the group of strongly connected anomalous users (Q3) called anomalous group.

Definition 2 (Anomalous Group and its Coreness): Let k be the largest degeneracy of Type-I network, G_1^T , then the connected k -core in the network is the anomalous group, and k is the coreness of the anomalous group in G_1^T . The coreness of an anomalous group $\mathcal{A}$ in G_x^T ($x = 2, 3$) is the minimum coreness of group members in G_x^T , which is denoted as $k_x^{\mathcal{A}}$.

Then, we provide the following measures.

Definition 3 (Common Neighbor Ratio and Diversity Ratio): Let $\mathcal{A}$ be an anomalous group, and $\mathcal{U}$ denote the set of other nodes in the connected $k_x^{\mathcal{A}}$ -core of network G_x^T . Define the common neighbor ratio (CNR) as,

$$r = \frac{\sum_{u \in \mathcal{U}} N_u}{|\mathcal{A}| |\mathcal{U}|}, \quad (3)$$

where N_u is the number of anomalous users in $\mathcal{A}$ who are connected to a user u . The diversity ratio (DR) is defined as,

$$\beta = \frac{|\{u | u \in \mathcal{U}, N_u \geq r \times |\mathcal{A}|\}|}{|\mathcal{A}|}. \quad (4)$$

The common neighbor ratio r can be interpreted as the fraction of anomalous users that interact with a user in the core on an average. A large r implies that these anomalous users interact with others collaboratively in this core. The diversity ratio β can be interpreted as the fraction of users that have interaction with the anomalous group. Fig. 3 illustrates three typical patterns. The first one shows a pattern with large r and small β ($r = 1$ and $\beta = 1/3$). For groups with this pattern, group members, i.e., anomalous users, have similar mention behavior and they only interact with a few users. The

second pattern has large r and large β ($r = 1$ and $\beta = 2$). For groups with this pattern, group members have similar behavior, in addition, they interact with a large number of users collaboratively. For the third pattern with small r and large β ($r = 1/3$ and $\beta = 2$), group members show diverse mention behaviors.

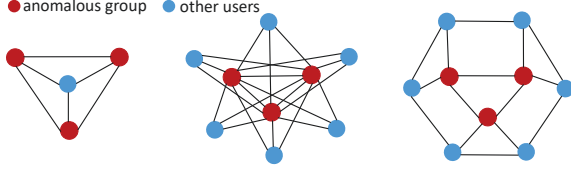


Fig. 3: Illustration of group behavior patterns.

V. EXPERIMENT RESULTS

The primary data set for this case study is a collection of “politics related” tweets obtained by using the Twitter streaming API, using a set of keywords as a filter. The filtering keywords are names of politicians (e.g., members of the US Congress) and topics at the center of public debates (healthcare, tax reform, Russia, abortion...). In order to eliminate sampling bias, we randomly sample from tweets after filtering. A majority of the collected data of 1.6 billion tweets spans from 2016 till date. In our experiment, we selected a 6-week time period from October 1 to November 14, 2018 as observation window, which is around the 2018 United States mid-term elections held on November 6, 2018. We break down the whole time interval to 6 short time intervals of 7 or 8 days each.

Anomalous users are selected out for each time period by following the framework illustrated in Algorithm 1. The two hyper parameters, δ and λ , are set to 1.5 and 1 respectively. Table I shows the fraction of ETT users and the fraction of anomalous users (AU) over ETT users during each time period.

TABLE I: Percentages of ETT users and Anomalous Users

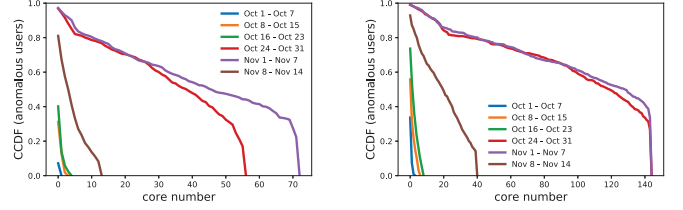
Time Period	Users Count	ETT (% of Users)	AU (% of ETT)
Oct 1 - 7	52,262	7.54	8.15
Oct 8 - 15	76,935	8.57	8.51
Oct 16 - 23	89,810	10.31	9.38
Oct 24 - 31	122,570	6.14	11.69
Nov 1 - 7	96,858	7.36	11.03
Nov 8 - 14	69,060	8.19	7.92

A. User Distribution over Coreness

To analyze the dynamic social connectivity between anomalous users and other user categories, we construct the three types of networks for each time period.

1) *Interaction among anomalous users*: We evaluate the distribution of anomalous users over coreness in Type-I social network at different time periods, and Fig. 4(a) shows the complementary cumulative distribution function (CCDF). In the first three weeks of October, most users do not interact with each other since more than half of anomalous users have 0 coreness. However, when it is near the American midterm election, i.e., from Oct 24 - Nov 7, the red and purple lines suggest that users concentrate on high coreness.

There are only around 700 anomalous users at Nov 1 - 7 from Table I, while the largest coreness is 72, which implies extremely strong connections amongst them. However, after the election, the largest coreness dropped to 13, which means that many anomalous users leave hot interaction with others. The distribution of users tends to be more uniform since the brown line decreases smoothly.



(a) Type I Network

(b) Type II Network

Fig. 4: CCDF of anomalous users over coreness.

2) *Interaction between anomalous users and other user types*: We evaluate the coreness distribution of anomalous users when their ETT- or regular neighbors are included. Fig. 5 shows the CCDF of anomalous users over coreness in three types of networks during Nov 1 - 7. When including first neighbors, the coreness of anomalous users increases dramatically, suggesting that many anomalous users have strong connections with their ETT neighbors or regular neighbors. The largest coreness increases from 72 to around 140 and 200. However, the number of anomalous users is only around 11% of ETTs and less than 0.8% of regular users at Nov 1 - 7. Considering the sizes of these three user categories, the increase of coreness of anomalous users is of less significance. Thus, we conclude that many anomalous users indeed have strong connections with some extreme tweeters or regular users, however, from a holistic view, interactions amongst anomalous users themselves are stronger than those between anomalous users with ETT or regular users.

To analyze the dynamic pattern of their interactions, we also plot the CCDF of anomalous users for Type-II and III networks and since they show the similar trend of changes over time, Fig. 4(b) only shows the plot for Type-II. Similar to the trend of Type-I network as we have analyzed, the interaction between anomalous users and other two types of users become much more intense during Oct 24 - Nov 7 and become sparse after midterm election (Nov 8 - 14).

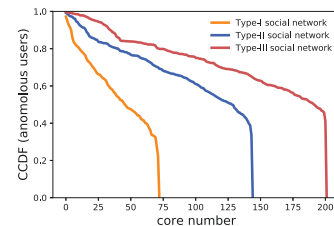


Fig. 5: CCDF of anomalous users in three networks during Nov 1 - 7

B. Anomalous Group Behavior Pattern

For first three weeks of our observation window, there is no anomalous group formed. We find an anomalous group in each of the other three time periods. Table II presents their group coreness in Type-I and II networks (denoted as Coreness 1 and Coreness 2), and their common neighbor ratio and diversity ratio in Type-II connection pattern.

TABLE II: Anomalous Groups

Time Period	Coreness 1	Coreness 2	CNR	DR
Oct 24 - 31	56	143	0.37	1.59
Nov 1 - 7	72	144	0.40	0.89
Nov 8 - 14	13	37	0.21	4.62

During Nov 1 - 7, as the relatively high CNR and low DR values suggest, the anomalous users in this group (Group I) tend to have similar mention behaviors and they do not interact with many people comparing to their group size. Instead, during Nov 8 - 14, anomalous users in the group (Group II) tend to have diverse behaviors, i.e., they do not interact with others collaboratively and they mention a lot of people.

C. Topic Analysis for Anomalous Groups

Group I and II also show completely different mention behaviors. We collect all hashtags used by members of each group to mention i) regular users (regular hashtags), ii) non-anomalous ETT (ETT hashtags) and iii) each other within group (group hashtags). Fig. 6 shows the count percentages of a same subset of hashtags for the two groups. We calculate

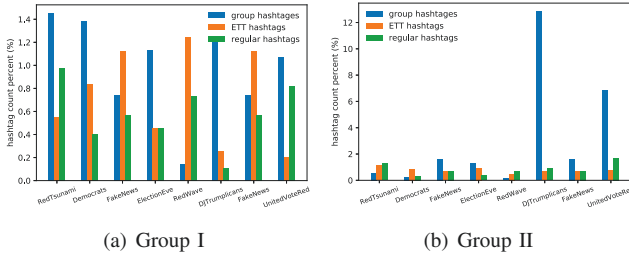


Fig. 6: Hashtag histogram for anomalous group.

the correlation coefficient between counts of group hashtags and ETT hashtags (coef1) and that between counts of group hashtags and regular hashtags (coef2) as shown in Table III. Besides, the standard deviations (stdev) for group hashtags, ETT hashtags and regular hashtags distributions (stdev1 , stdev2 , stdev3 respectively in Table III) are calculated for each group. Note that the stdev presented in this section is standardized by the percentage of hashtag counts.

TABLE III: Statistics of Hashtag Distributions

	coef1	coef2	stdev1	stdev2	stdev3
Group I	0.81	0.77	0.004	0.003	0.002
Group II	0.31	0.10	0.024	0.008	0.004

For Group I, it has similar intra-group and inter-group hashtag usage behaviors as two correlation coefficients indicate, and all the three distributions have very low stdev , indicating a uniform use of hashtags. For Group II, however, the low coefficients suggest that this group has different hashtag uses when they mention amongst group members versus mention

other categories. In addition, the distribution of group hashtags (blue bars in Fig. 6) has much higher stdev than the other two distributions (yellow and green bars in Fig. 6). It indicates that some hashtags (e.g., DJTrumpicans or UnitedVoteRed) are used disproportionately often when members of Group II interact amongst themselves. However, when they mention other categories of users, their hashtag distribution is more uniform, i.e., the conversation covers broader topics.

VI. CONCLUSION

Our user stratification strategy successfully brings out the unusual behavior of the extreme tweeters. The strategy itself is fairly generic and can be applied to any social media platform. We find that the highly-connected group we call “anomalous” exhibits a “clannish” behavior with strong and narrowly-scoped within-group interactions that markedly differs from their across-group behavior. It suggests that the group has strong political beliefs and forms a strong trust network that establishes contact with a high number of users but carefully controls content outside the group. Our future work will elaborate this case study to uncover other characteristics of these user groups and their interactions.

Acknowledgment. This work was partially funded by the National Science Foundation grant (Number 1738411).

REFERENCES

- [1] H. Shaban, “Twitter reveals its daily active user numbers for the first time,” *Washington Post*, Feb. 2019.
- [2] X. Zheng and A. Gupta, “Social network of extreme tweeters: A case study,” *arXiv preprint arXiv:1905.00567*, 2019.
- [3] T. Tuna, E. Akbas, A. Aksoy, M. A. Canbaz, U. Karabiyik, B. Gonen, and R. Aygun, “User characterization for online social networks,” *Social Network Analysis and Mining*, vol. 6, no. 1, p. 104, 2016.
- [4] P. A. Grabowicz, L. M. Aiello, V. M. Eguiluz, and A. Jaimies, “Distinguishing topical and social groups based on common identity and bond theory,” in *Proc. of the 6th ACM Int. Conf. on Web search and data mining*. ACM, 2013, pp. 627–636.
- [5] J. Jang and S.-H. Myaeng, “Discovering dedicators with topic-based semantic social networks,” in *Proc. of the 7th Int. AAAI Conf. on Weblogs and Social Media*, 2013.
- [6] F. T. O’Donovan, C. Fournelle, S. Gaffigan, O. Brdiczka, J. Shen, J. Liu, and K. E. Moore, “Characterizing user behavior and information propagation on a social multimedia network,” in *2013 IEEE Int. Conf. on Multimedia and Expo Workshops*. IEEE, 2013, pp. 1–6.
- [7] C. A. Bail, “Combining natural language processing and network analysis to examine how advocacy organizations stimulate conversation on social media,” *Proc. of the National Academy of Sciences*, vol. 113, no. 42, pp. 11 823–11 828, 2016.
- [8] N. Halko, P.-G. Martinsson, and J. A. Tropp, “Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions,” *SIAM review*, vol. 53, no. 2, pp. 217–288, 2011.
- [9] E. J. Candès and B. Recht, “Exact matrix completion via convex optimization,” *Found. of Comput. Mathematics*, vol. 9, no. 6, p. 717, 2009.
- [10] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, “K-core organization of complex networks,” *Phys. Rev. Letters*, vol. 96, no. 4, p. 40601, 2006.
- [11] J. I. Alvarez-Hamelin, L. Dall’Asta, A. Barrat, and A. Vespignani, “Large scale networks fingerprinting and visualization using the k-core decomposition,” in *Adv. in Neural Info. Proc. Syst.*, 2006, pp. 41–50.
- [12] J. Ugander, L. Backstrom, C. Marlow, and J. Kleinberg, “Structural diversity in social contagion,” *Proc. of the National Academy of Sciences*, vol. 109, no. 16, pp. 5962–5966, 2012.
- [13] Q. Huang, V. K. Singh, and P. K. Atrey, “Cyber bullying detection using social and textual analysis,” in *Proc. of the 3rd Int. Workshop on Socially-Aware Multimedia*. ACM, 2014, pp. 3–6.