Noname manuscript No.
(will be inserted by the editor)

Examining the Evolution of the Twitter Elite Network

Reza Motamedi · Soheil Jamshidi · Reza Rejaie · Walter Willinger

the date of receipt and acceptance should be inserted later

Abstract The most-followed Twitter users and their pairwise relationships form a sub-graph of Twitter users that we call the *Twitter elite network*. The connectivity patterns and information exchanges (in terms of replies and retweets) among these elite users illustrate how the "important" users connect and interact with one another on Twitter. At the same time, such an elite-focused view also provides valuable information about the structure of the Twitter network as a whole.

This paper presents a detailed characterization of the structure and evolution of the top-10K Twitter elite network. We describe our technique for efficiently and accurately constructing the Twitter elite network along with social attributes of individual elite accounts and apply it to capture two snapshots of the top-10K elite network that are some 2.75 years apart. We show that a sufficiently large elite network is typically composed of 14-20 stable and cohesive communities that are recognizable in both snapshots, thus representing "socially meaningful" components of the elite network. We examine the changes in the identity and connectivity of individual elite users over time and characterize the community-level structure of the elite network in terms of bias in directed pairwise connectivity and relative reachability. We also show that both the reply and retweet activity between elite users are effectively contained within individual elite communities and are generally aligned with the centrality of the elite community users in both snapshots of the elite network. Finally, we observe that the majority of the regular Twitter users tend to have elite friends that belong to a single elite community. This finding offers a promising criterion for grouping regular users into "shadow partitions" based on their association with elite communities.

University of Oregon · NIKSUN, Inc. E-mail: {motamedi, jamshidi, reza}@cs.uoregon.edu · wwillinger@niksun.com

1 Introduction

The increasing popularity of online social networks (OSN) such as Twitter and Facebook has fueled a growing interest in studying their connectivity structure and its evolution over time (*e.g.*, [11]), estimation of their user population (*e.g.*, [18]), migration of users between different OSNs (*e.g.*, [22]), the exchange of information among their users (*e.g.*, [23]), and how their users influence each other (*e.g.*, [7]). These studies usually focus on the entire social network (*e.g.*, the entire connectivity structure [4, 14] or all users). However, a majority of OSN users typically have a low level of connectivity and/or activity (*e.g.*, have only a handful of followers or tweet just a few times a month). Therefore, any analysis of the entire network tends to be dominated by these regular users.

We argue that high-degree and/or highly-active or influential (or "elite") users [1] in an OSN play a more important role in terms of connectivity, information propagation and influence than regular ("non-elite") users. Therefore, characterizing the connectivity structure of and the information exchange within the *elite network*, the core subgraph of an OSN that contains all the "elite" users and their pairwise relationships promises a number of interesting opportunities. First, characterizing the elite network has the potential of revealing the relationship and influence patterns between the elite users in a particular OSN. Such findings can provide valuable insights into how the OSN is used by these important users. Second, elite users are usually well-known individuals/entities (*e.g.*, celebrities, news agencies, politicians, organizations) with specific social, cultural, or geographic attributes that can be leveraged to examine relationship patterns among elite users in a *socially-informed* manner. Lastly, since elite users collectively have direct connectivity to a significant number of regular users in an OSN, identifying certain relationship patterns among elite users

has the potential of providing valuable insights into the overall structure of the entire OSN network. The key challenge in characterizing the elite network of an OSN is to accurately capture the elite network at the "right" level of granularity (*e.g.*, type of users and type of activities). To our knowledge, no prior study has tackled the problem of capturing and characterizing elite network in Twitter and other major OSNs.

Motivated by these observations, this paper presents the first detailed characterization of the *Twitter elite network* and its evolution over time. To this end, we first describe our proposed methodology for accurately and efficiently capturing two snapshots of the Twitter elite network in January of 2016 and October of 2018 that consist of the top 10K most-followed Twitter users (10K-ELITE) at those times, along with their social attributes and their pairwise follower-friend relationships. We consider the resulting two elite networks at different levels of granularity or size (*i.e.*, view) and study how their overall connectivity structure compares across different views. Focusing on a particular view (*i.e.*, 10K-ELITE), we examine some basic characteristics of the two snapshots of Twitter's elite network and show how the connectivity of individual elite users changes within the elite network and the entire Twitter network from one snapshot to the next. Next, we identify resilient "elite" communities in the captured snapshots of our elite networks and explore the social attributes of the elite users in each of these communities to determine whether they exhibit any specific theme. We illustrate how elite communities evolve over time and how they grow, split, and merge as we consider the elite networks at finer levels of granularity (*i.e.*, larger sizes), from 1K to 10K nodes. The identified elite communities enable us to examine the structure of the Twitter elite network at the community level. More specifically, to study the pairwise connectivity between elite communities in each snapshot along with their temporal evolution, we consider two metrics: the level of bias in directed edges between communities and the pairwise reachability between communities. We also utilize new metrics to assess the retweet (and reply) influence of elites at the user-level and community-level in the elite network. Furthermore, we investigate the alignment between these measures of influence and the "importance" of a user's location in the elite network as measured by PageRank [6]. Finally, we briefly explore the relationship between regular and elite users to determine whether elite communities can be leveraged to cluster regular Twitter users in a meaningful and cohesive manner. This study is an extension of our earlier work on characterizing (a single snapshot of) the Twitter elite network [16]. In particular, by considering two snapshots of the Twitter elite network taken some 2.75 years apart, this paper provides a first look at the temporal evolution of the Twitter elite network and of some of its main characteristics.

Our characterization of the structure and evolution of the Twitter elite network results in a number of specific findings. First, irrespective of the used cut-off (*e.g.*, top 10K most-followed users), all nodes in a Twitter elite network form a weakly connected component that has a star-shaped structure in both snapshots. The largest strongly connected component (LSCC) that contains a vast majority of all the nodes and edges is at the center of this structure and is typically surrounded by peripheral nodes that have directed edges to (*i.e.*, being followed by users in) the LSCC. While elite users become in general more popular among regular users over time, the density of the elite network shrinks over time. Second, the Twitter 10K-ELITE network is composed of some 14 to 20 *elite communities* of different sizes that exhibit "social cohesion" around a common theme related to, for example, a country, a language, a certain cultural background, or a particular business interest. The number of elite communities increases and their identities further diversify over time. Within each snapshot, the number of elite communities and their associated themes remain rather stable once the elite network reaches a certain size (i.e., some 6K nodes). Furthermore, most communities remain recognizable with a similar theme across different snapshots of the elite network. These observations suggest that elite communities define robust and socially meaningful entities of the Twitter elite network. Third, examining the resulting elite communities in more detail, we observe a symmetric negative bias in the directed connectivity between a few elite communities and the rest of the elite communities. We observe that elite communities form a few groups with significantly higher pairwise reachability within each group. However, these groups of more reachable communities have typically evolved over time. This higher reachability can be explained by a subset of elite users that are not part of any elite community but act as "bridges" between different elite communities. Fourth, to our surprise, we find that for all elite communities, the normalized retweet and reply influence of individual elite users is primarily contained to other users in the same community. We also show that both the aggregate level of influence and the number of influenced users should be considered as two separate dimensions of user-level retweet (or reply) influence. While most of the influential users remain stable between both snapshot, we also observe new influential users appearing in the second snapshot (*e.g.*,, Donald Trump appears as a new user with a dominant retweet and reply influence in 2018). We show that for both snapshots, the retweet and reply influence of users in most elite communities is proportional to (and thus likely determined by) their ranking based on the centrality in the elite network. However, a few particular communities (*e.g.*, Adult) do not exhibit this characteristic, and their users have clearly a higher (or lower) level of influence than what their rankings indicate.

Finally, we observe that a majority of the elite friends of regular Twitter users tend to belong to a single elite community. This finding suggests a promising criterion for grouping regular users into "shadow partitions" based on their association with users in elite communities. In particular, we show that the overall inter-connectivity between these shadow partitions closely mirrors the inter-connectivity between the corresponding elite communities which, in turn, suggests that the identified shadow partitions can be viewed as extensions of their corresponding elite communities.

The rest of the paper is organized as follows. In Section 2, we present our technique for capturing our snapshots of the Twitter elite network. We explore the effect of the size of the elite network in Section 3 and present basic characteristics of the two snapshots of the Twitter elite network along with their changes over time in Section 4. Our approach for detecting elite communities, identifying their basic characteristics and studying their evolution over time is described in Section 5. Aspects of the inter-connectivity and cross-influence among elite communities for both snapshots are discussed in Section 6 and 7, respectively. In Section 8, we explain how the association of regular Twitter users with individual elite communities can be leveraged to group those regular users into shadow partitions. We conclude in Section 9 by summarizing our contributions and outlining our plans for future work on this topic.

2 Capturing the Twitter Elite Network

Our goal is to efficiently capture the Twitter elite network; that is a subgraph of the Twitter OSN that contains the top-N most-followed accounts (*i.e.*, nodes) and the friend-follower[1] relationships among them (*i.e.*, edges)[2]. Furthermore, we want to annotate each node with its social and geographical (location) attributes and use this such annotated graph as input for our analysis. Our data collection strategy for capturing the Twitter elite network consists of the following four steps: *(i)* Collecting a list of the most-followed Twitter accounts using online resources and complementing that with additional accounts obtained by performing customized random walks, *(ii)* Identifying the pairwise connections between these accounts, *(iii)* Detecting any missing elite accounts and collecting their information, and *(iv)* Collecting all profile information and available tweets associated with the collected elite accounts. Next, we describe each of these four steps in more detail.

Step 1: To bootstrap the data collection process, we crawl lists of the most followed accounts from online resources. In particular, marketing websites such as `socialbakers.com` offer professionally maintained lists of the most followed accounts in a variety of OSNs in different social categories (*e.g.*, celebrities, actors, sport, community, ...). For example, each list on `socialbakers.com` provides up to $1\,000$ top accounts in the selected category along with the number of followers and username for each account. We collect the list associated with all offered categories and subcategories and create a unified list that includes all the uniquely-discovered user accounts with their number of followers (and associated rank), their category and location from `socialbakers.com`. This resulting unified list consists of $59,832$ unique users whose number of followers varies from 263 to 81M, and they are associated with 123 categories and 191 unique countries.

To independently identify Twitter accounts with many followers, we also conduct 2K "customized" random walks over the Twitter graph that starts from randomly selected Twitter accounts. Our random walks only select a random user from the friend list of the current user as their next step. The likelihood that these walkers visit a user is proportional to its number of followers. Therefore, these random walks offer an efficient technique to identify the most-followed and visible users[3] [21]. We merge all the discovered accounts from our random walks with the accounts captured from `socialbakers.com` into a single sorted list based on the number of followers and only focus on the top 10K accounts with the most followers to form a *master list*.

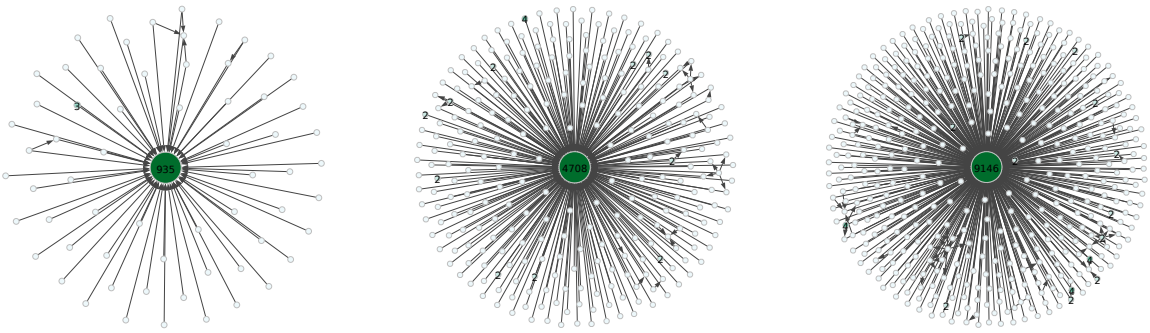
Step 2: The relationship between elite user u and its follower v (*i.e.*, directed edge $u \rightarrow v$ in the elite network) can be discovered from either end. More specifically, we can *(i)* obtain the list of all u's followers and discover follower v among them, or *(ii)* collect the list of all v's friends and discover friend u among them. Our observation is that the number of friends of elite accounts is typically several orders of magnitude smaller than the number of followers. Since the overhead of data collection is directly proportional to the length collected list, we adopt the second crawling strategy and discover all edges by obtaining the list of friends of elite users. Using this strategy, the total number of crawled friend-follower relationships is 504.8M and results in 95M unique friends for the top 10K most-followed elites.

Step 3: At this point, we have a snapshot of the directed Twitter subgraph that connects the most-followed Twitter accounts. Since it is possible that the identified top-10K accounts in step 1 do not accurately represent the actual top-10K accounts on Twitter, we perform one more check to examine whether the list of identified account is correct and complete. We observe that any missing elite account is very likely to be followed by many elites that we already identi-

[1] In Twitter, each user u has a collection of *followers* that receive any tweets that u sends. u is called a *friend* for each one of its followers.

[2] We use the terms *nodes with the highest degree* and the *most-followed accounts* interchangeably.

[3] A user with many followers that is part of a partition or weakly connected region is not likely to be discovered by random walks. We argue that such an elite user is less important for our analysis.



(a) 1K-ELITE (b) 5K-ELITE (c) 10K-ELITE

Fig. 1 The connectivity of strongly connected components of the elite networks.

fied as top 10K accounts [2]. Since we already collected the entire list of friends for the top-10K accounts, we can calculate the number of elite-followers for all these collected friends that are not among the elites, and sort the resulting list by the number of elite-followers. We start by scanning this list from the top and collect account information (*e.g.*, number of followers) for users in this list. If the number of followers for any of these accounts is larger than the number of followers for the account at rank 10K in our current list, we add it to the master list (at the proper rank) and update the ranks for all elites. We continue this process until 100 consecutive accounts from this sorted list do not make it to the master list. We finally identify the edges between these newly added accounts and other top 10K accounts by collecting their friend lists. Using this technique, we detected less than 3% missing elite accounts that are between ranks 500 and 10K. The small percentage of the discovered missing accounts in this step along with their relatively low ranking indicates that for all practical purposes, our master list can be considered to be accurate and complete. *In summary, among the top* 10*K most-followed Twitter accounts, roughly 87% are exclusively reported in* `socialbakers.com`*, 3% are found exclusively using our random walk method, 7% are confirmed by both of the above two techniques, and 3% are among the discovered friends of most-followed accounts.*

Step 4: We collect all the available tweets (up to the last $3,200$ which is the maximum number provided by Twitter) for each top 10K Twitter account. These tweets are used to investigate the influence of elites by analyzing retweets and also to gain insight into how the top 10K elites use Twitter by analyzing tweets/retweets and constructing topic models.

Datasets: Using the described 4-step process, we captured a snapshot of the 10K-ELITE network of Twitter in January of 2016 and the second snapshot in October of 2018. We refer to these two snapshots as $S16$ and $S18$. There are 7100 overlapping elites between these two snapshots. These two datasets enable us to study how the composition of the Twitter elite network, its connectivity structure and the influence among elite users have evolved over a period of 33 months.

3 Different Views of the Twitter Elite Network

While it is compelling to consider Twitter users with the highest number of followers as Twitter elites, one remaining question is *how many of the most-followed accounts should be considered as a part of the elite network?* We argue that the 10K-ELITE offers a sufficiently detailed view of Twitter's elite network. For one, the skewed distribution of the number of followers implies that the number of followers rapidly drops with rank. For example, in our datasets, the top 10 most-followed accounts have between 43.9 and 81.7M followers in $S16$ (with a median of 0.93M) and between 59.4 and 107.8M follower in $S18$ (with a median of 1.4M). In contrast, the last 10 accounts in the top 10K have around 0.4M followers in $S16$ and some 0.7M in $S18$. Therefore, the popularity (and thus importance) of any account beyond the top 10K would be significantly lower. Second, examining the friend lists of 10K randomly selected Twitter users reveals that 80% of these random (and thus 80% of all) Twitter accounts follow the top 10K elite accounts. Third, while it is feasible to capture a larger elite network beyond 10K, reliably collecting the desired attributes (*e.g.*, social and location information) for these users is very expensive and their addition has a diminishing rate of return. To examine *whether and how the size of the resulting elite network affects its structural properties*, we consider the Twitter elite network at different sizes or levels of granularity (*i.e.*, views). Each view, referred to as nK-ELITE, contains the top n*-thousand* most-followed accounts and friend-follower relationships between them. Interestingly, we observe that all considered views of the elite network con-

Table 1 Basic connectivity features of different views of the elite network and their strongly connected components for $S16$ and $S18$ napshots

	ALL_2016				LSCC_2016		All_2018				LSCC_2018	
View	$\|E\|$	Rcp	Diam	#SCC	$\%\|V\|$	$\%\|E\|$	$\|E\|$	Rcp	Diam	#SCC	$\%\|V\|$	$\%\|E\|$
1K-Elites	40K	0.35	4	64	93.5	94.55	37K	0.38	5	70	91.89	94.17
2K-Elites	104K	0.34	3	109	94.25	95.61	100K	0.36	4	139	92.74	94.81
3K-Elites	194K	0.32	4	169	94.13	95.76	181K	0.34	4	215	92.6	94.97
4K-Elites	289K	0.31	4	229	94.02	95.95	273K	0.34	4	264	93.15	95.36
5K-Elites	413K	0.32	4	277	94.2	96.13	376K	0.33	4	301	93.76	95.65
6K-Elites	543K	0.33	4	336	94.12	96.21	485K	0.33	4	355	93.86	95.77
7K-Elites	676K	0.34	4	369	94.5	96.44	607K	0.34	4	409	93.94	95.86
8K-Elites	841K	0.37	4	400	94.77	96.66	735K	0.34	4	454	94.11	95.96
9K-Elites	1.01M	0.4	4	438	94.92	96.91	903K	0.34	4	484	94.43	96.14
10K-Elites	1.15M	0.42	4	453	95.08	97.05	1.1M	0.35	4	526	94.55	96.23

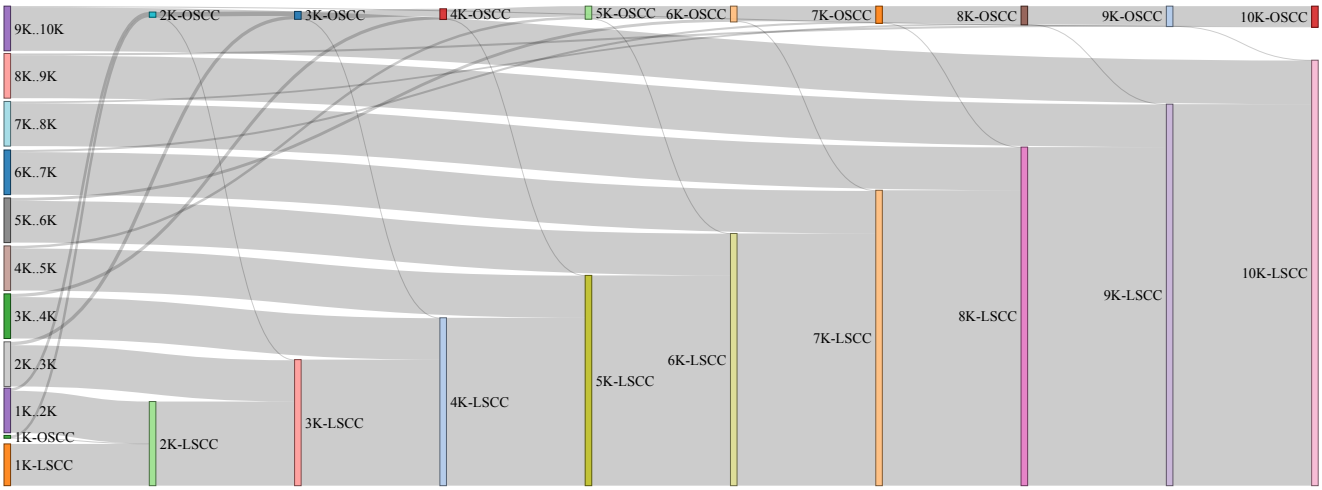


Fig. 2 The dynamics of mapping of elite users between the LSCC and other SCCs OSCC in each view as the size of elite network increases from 1K to 10K nodes.

sist of a single weakly connected component that contains all top *n-thousand* nodes. Furthermore, the largest strongly connected component (LSCC) [9] in each view contains 91-95% of all nodes and 94-97% of all edges in the elite network. While most of the other strongly connected components (SCCs) consist of a single node, we typically observe only a few SCCs that have two or more nodes. Figure 1 visualizes the strongly connected component structure of 1K-ELITE, 5K-ELITE, and 10K-ELITE as directed graphs where each circle represents a SCC with the number indicating the number of nodes in that SCC. LSCC is shown with a green circle in the center. Arrows represent friend→follower relationships between users in different SCCs. This figure shows that across the different views, SCCs form a "star-like" structure where the LSCC is in the center and there are directed edges from other SCCs to nodes in the LSCC (*i.e.*, only elite users in the LSCC follow and receive tweets from elite users in other SCCs).

As more nodes are included in the view of the elite network, other SCCs in one view may be pulled into the LSCC in the next view since the extended view may include more shortcuts. Figure 2 illustrates via a Sankey diagram [24] how nodes in the LSCC and other SCCs of each view are mapped to the LSCC and SCCs in the next view. In this figure, individual views (from 1K to 10K) are shown along the x-axis. For each view, the vertical box at the bottom of each column (labeled as *K-LSCC) represents the LSCC and the box on the very top represents other SCCs (labeled as *K-OSCC). Groups of elites ranked by their number of followers are all presented in the first column alongside 1K-ELITE and labeled as 1K..2K, 2K..3K, etc. Extending the elite network adds one of the groups to a view to create the next view. For instance, elites with rank [1K...2K] join 1K-ELITE to create 2K-ELITE. As the plot shows, in each view more than 95% of these newly added elites join the LSCC and the remaining elites join the other SCCs. A close examination of these views also reveals that roughly 13-20% of nodes in OSCC are pulled into the LSCC in the next view when one of their friends becomes part of the elite network. Note however that the group of other SCCs has no friends (*i.e.*, no incoming edges) and thus remains outside the LSCC regardless of the size of the elite network.

Table 2 The node level connectivity of the top 10 most-followed elites in 2018 across both snapshots

name	screen_name	followers_2018	friends_2018	followers_2016	friends_2016	Rank_change	Cur._Rank
KATY PERRY	katyperry	107.8M	<1K	81.8M	<1K	0	1
Justin Bieber	justinbieber	105.2M	303K	74.3M	254K	0	2
Barack Obama	BarackObama	103.5M	618K	69.1M	638K	-1	3
Rihanna	rihanna	88.7M	1K	55.4M	1K	-2	4
Taylor Swift	taylorswift13	84.0M	<1K	70.2M	<1K	2	5
Lady Gaga	ladygaga	77.8M	126K	55.0M	131K	-1	6
Ellen DeGeneres	TheEllenShow	77.1M	36K	53.3M	37K	-1	7
Cristiano Ronaldo	Cristiano	75.7M	<1K	40.2M	<1K	-6	8
YouTube	YouTube	71.6M	1K	59.0M	1K	4	9
Justin Timberlake	jtimberlake	65.1M	<1K	51.6M	<1K	0	10

4 A First Look at the Twitter Elite Network

Before conducting any detailed analysis of the Twitter elite network, we first describe some of its basic macro-level characteristics and present some observations of the evolution of its node-level connectivity.

4.1 Macro-Level Connectivity Features

Table 1 lists for both the $S16$ and $S18$ snapshots a number of basic characteristics for each view of their elite network, including the number of directed edges ($|E|$), its reciprocity (Rcp), the diameter (Diam), and the number of SCCs (#SCC). This table clearly shows that as the size of the elite network grows (from 1K to 10K), it becomes denser (the average degree obtained by dividing the number of directed edged $|E|$ by the number of nodes given by the size of the elite network or view in the first column increases from roughly 40 to 115 for $S16$ and 37 to 110 for $S18$, the fraction of reciprocated edges initially drops and then increases, and its diameter remains stable. In all views, roughly 31-42% of the friend-follower relationships are reciprocal, which is higher compared to the reported 22% for the entire Twitter social graph [12]. For each snapshot, Table 1 also presents the fraction of all nodes ($\%|V|$) and edges ($\%|E|$) of each view that are part of its LSCC. We observe that the LSCC in each view contains 91-95% of all nodes and 94-97% of all edges of the corresponding elite network.

4.2 Evolution of Node-Level Connectivity

We first examine how the micro-level or node-level connectivity of the elite network has evolved over the 33-months long period between our two snapshots. A comparison of the 10K-ELITE network in the $S16$ and $S18$ snapshots shows that 81% of elite users and 84% of edges (relationships) between elites are common between them. This suggests that a majority of nodes and edges in 10K-ELITE network have remained stable. However, upon closer inspection of the number of edges and the percentage of reciprocal edges in each snapshot (*i.e.*, $|E|$ and Rcp columns in Table 1), we observe that both the density and level of reciprocity of the elite network have decreased in all views of the elite network.

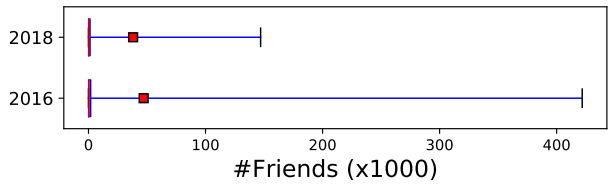


Fig. 3 Distribution of total number of friends among elites in both snapshots

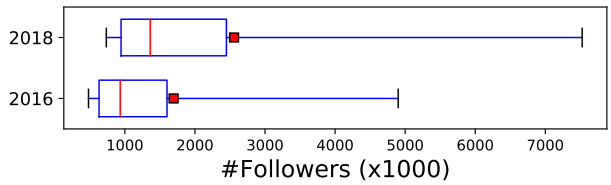


Fig. 4 Distribution of total number of followers among elites in both snapshots

To examine in more detail how the overall connectivity of individual elite users has evolved over time, Table 2 includes the number of followers and friends for the top 10 most-followed elite users in $S18$ in both snapshots. The table shows that while the total number of followers for all these users has significantly increased, their overall ranking among elites (shown in the last column) experiences no or only a small change. Figures 3 and 4 present the summary distribution of the number of all (*i.e.*, elite and regular) followers and friends among elite users in both snapshots. These figures show that the number of all friends for elite users is typically small and even slightly decreased between the two snapshots. In contrast, the number of followers is typically a couple of millions, exhibits a skewed distribution and has further increased (and become more skewed) between the two snapshots. We further explore the changes in the connectivity of individual nodes. Figures 5 and 6 depict

Table 3 Users among top 100 elites in $S16$ that experienced the most positive rank improvement between the two snapshots

name	screen_name	followers_2018	friends_2018	followers_2016	friends_2016	Rank_change	Current_Rank
Elon Musk	elonmusk	23.4M	<1K	3.3M	<1K	827	83
Hillary Clinton	HillaryClinton	23.6M	1K	5.2M	1K	341	80
Donald J. Trump	realDonaldTrump	55.4M	<1K	5.9M	<1K	341	17
Shawn Mendes	ShawnMendes	20.2M	60K	5.5M	49K	298	96
Virat Kohli	imVkohli	27.1M	<1K	9.2M	<1K	106	64

Table 4 Users among top 100 elites in $S16$ that experienced the most rank drop between the two snapshots

name	screen_name	followers_2018	friends_2018	followers_2016	friends_2016	Rank_change	Current_Rank
Avril Lavigne	AvrilLavigne	22.0M	<1K	19.7M	<1K	-36	87
David Guetta	davidguetta	21.7M	<1K	19.4M	<1K	-35	88
Marshall Mathers	Eminem	22.8M	<1K	20.0M	<1K	-34	84
Ed Sheeran	edsheeran	19.6M	<1K	16.6M	1K	-33	98
Adele	Adele	28.3M	<1K	24.9M	<1K	-27	58

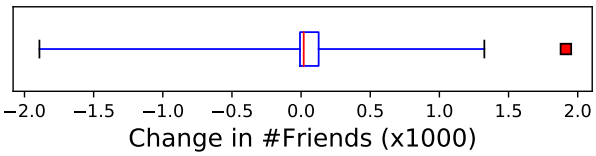


Fig. 5 Summary distribution of changes in # Friends over two years

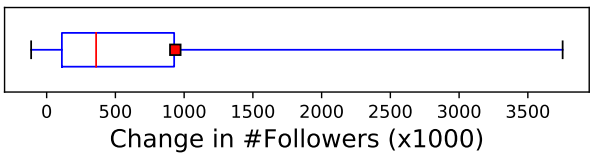


Fig. 6 Summary distribution of changes in # Followers over two years

the summary distribution of changes in the number of followers and friends across 7100 overlapping elites for the two snapshots. These plots clearly indicate that for a significant majority of nodes, the total number of followers and friends have indeed increased. To assess the effect of changes in the total number of followers on the relative ranking of elite users, Figure 7 presents the changes on the ranking across the overlapping elite users in both snapshots. We notice that the ranking for 67% of these elite users improve between 0-4000 while the ranking for 32% of the elite users dropped within the same range.

Tables 3 and 4 present the node level connectivity for five users among the top-100 elites in 2018 that experience the largest improvement or drop in ranking between the two snapshots, respectively. These examples also illustrate the types of elite users over time. Interestingly, the rank of all five accounts in Table 4 has dropped despite a significant increase in their total number of followers.

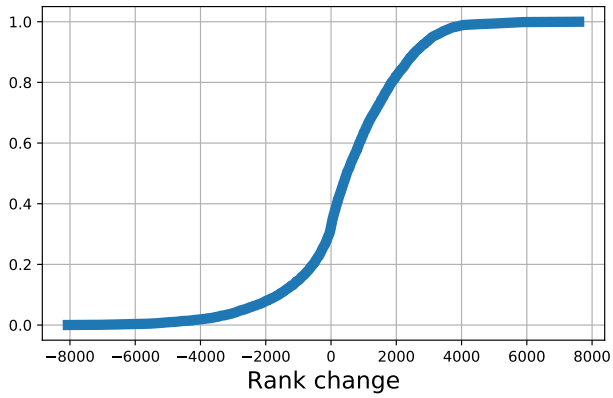


Fig. 7 CDF of changes in elite ranking between two snapshots (positive changes means improvement in ranking)

*In summary, our results show that the elite users experienced an increase in the total number of followers between the two snapshots (*i.e.*, elite users have become more popular over time as shown in Figure 4) while the density of the elite network (average number of connection to other elite users) has indeed decreased (*i.e.*, $\frac{|E|}{|V|}$ in Table 1). In other words, elite users become more popular among regular users but less connected to each other between 2016 and 2018.*

5 Elite Communities

In this section, we examine whether the Twitter elite network is composed of a collection of meaningful components that can be used to gain insight into the relationship among Twitter's elite users as well as the overall structure of Twitter.

Label	Size	Dens.	Theme
US/PoP	2.9K	384	US celebs/actor/music
Spanish	1.9K	208	Spanish Speaking
US/Corp	1.3K	242	US Corporate/Media
Arabic	1K	698	Arabic Speaking
ID	533	93	Indonesian
BR	508	162	Brazilian
PH	475	210	Filipino
IN	335	185	Indian
TR	271	87	Turkish
Unstable	155	268	Unstable nodes
K/Pop	150	51	Korean Popstars
TH	28	34	Thai
Adult	20	57	Adult/Porn
US/TV	19	541	US TV channels
GLB/Fun	13	119	Global Entrtnmnt

Table 5 Label and key features of 14 elite communities in the 10K-ELITE Twitter elite network in the 2016 snapshot

Label	Size	Dens.	Theme
US/Pop	2K	279	US celebs/actor/music
Arabic	1.4K	180	Arabic Speaking
Spanish	1.3K	135	Spanish Speaking
US-Pol/Corp	1.1K	140	US politicians/News/Corp
IN	740	105	Indian
BR	530	116	Brazilian
US/SP	525	113	US sports (News/Corp)
GB/Soccer	449	104	Accounts related to soccer
ID	419	70	Indonesian
K/Pop	399	32	Korean Artists
Unstable	345	21	Unstable nodes
Entertain	207	65	Gaming Consoles/Game makers
PH	200	90	Filipino
FR	101	45	French Artists
PK	75	53	Pakistani
IT	54	39	Italian Artists
FR/Pol	53	45	French Politicians/News
RU	28	12	Russian Media/Artists
TH	15	8	Thai
Adult	14	20	Adult/Porn

Table 6 Label and key features of 20 elite communities in the 10K-ELITE Twitter elite network in the 2018 snapshot

5.1 Identifying Resilient Elite Communities

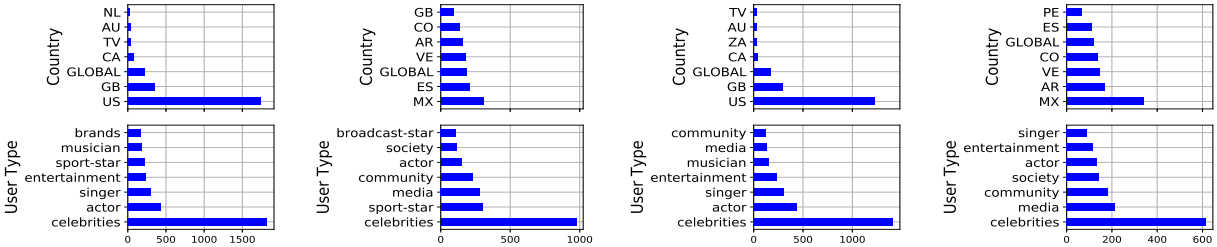
While it is tempting to assume that the most natural components of an OSN are groups of tightly connected nodes (or "communities"), applying out-of-box community detection algorithms is problematic. For one, while most commonly-used community detection techniques take as their input undirected graphs, the Twitter elite network is a directed graph [10]. To address this issue, we first convert each view of the elite network into a weighted undirected graph by replacing *each* directed edge into a single undirected edge, either with a weight of 2 when reciprocal directed edges exist or weight of 1 otherwise. Compared to prior studies (*e.g.*, [13] simply converts a directed graph into an undirected one), this representation allows us to encode tighter bindings between users with reciprocal edges. Second, the outcome of some of the most commonly-used community detection techniques (*e.g.*, Louvain [4], BigCalmm [25], InfoMap [19], EDA [17]) is non-deterministic and varies across multiple runs. To deal with this problem, we use the COMBO community detection technique [20] that formulates modularity optimization as on multi-objective optimization and therefore detects more stable communities across different runs as compared to, say, Louvain [4]. We also eliminate the residual instability by only considering a group of nodes as a community if these nodes were consistently grouped together across different runs. To achieve this objective, we run COMBO on each view of the elite network k times and collect the communities that individual nodes are mapped to in each run in vectors with k values, the so-called "community vectors". Then, we group all the nodes that are mapped to the same community in all runs (*i.e.*, they have the same community vector) and refer to the group as a *resilient community*. This process of detecting resilient communities also results in a group of nodes with unique community vectors. We group this set of nodes together with the nodes in resilient communities of a size smaller than 10 and refer to the resulting group as the set of *unstable nodes*.

Clearly, increasing k is more restrictive and may lead to smaller resilient communities since more runs can simply split a community into two (or more) smaller ones. We conservatively consider $k = 100$ in our analysis, as our empirical findings show that having more runs does not lead to the identification of more resilient communities in the elite network.

Running COMBO, we typically detect between 10-29 resilient communities across different views of each snapshot of the Twitter elite network. Collectively, these resilient communities cover 92-99% of all nodes in each view. *Thus, less than* 8% *of the elite users are* unstable *nodes in each snapshot.* We emphasize that the identified elite communities are different from the commonly-considered communities that one obtains by running community detection on the *entire* Twitter graph that contain many regular (*i.e.*, non-elite) users.

5.2 Elite Communities as Cohesive Entities

An important question is *whether the identified elite communities represent meaningful units of the Twitter network?* We answer this question by exploring whether users in each community exhibit social cohesion. Recall that `socialbakers.com` provides 8 social categories (and 137 subcategories) as well as 196 unique countries as the location attribute for more than 90% of elite users. Using this information, we examine the histogram of the social and geographic attributes (*i.e.*, footprint) across users in each elite community to assess their level of social "togetherness". Figure 8 shows the different footprints of two elite communities in



(a) US/Pop (in $S16$) (b) Spanish (in $S16$) (c) US/Pop (in $S18$) (d) Spanish (in $S18$)

Fig. 8 The social and geographic footprints for two sample elite communities in 10K-ELITE for both the $S16$ and $S18$ snapshots.

the 10K-ELITE view from each snapshot. A careful examination of these footprints shows that all of them exhibit a significant level of social and/or geographic (or language-based) cohesion. Since many elite accounts belong to easily recognizable individuals/entities, we manually inspect accounts in each community and leverage their social context to identify the "theme" associated with each community. Tables 5 and 6 summarize the main features of all the elite communities (that have more than 10 nodes) in 10K-ELITE for the $S16$ and $S18$ snapshots, respectively, and list the communities' assigned label, their size, their density, and their manually inferred theme [5]. While the level of cohesion varies among communities, all the communities exhibit a very pronounced theme. The themes can be broadly divided into the following four categories: *(i)* Elites from a single country (ID, BR, PH, TR, TH, PK, IT, RU, IN, ID); *(ii)* Elites from different countries with a common language (Spanish, Arabic); *(iii)* Elites with a similar cultural interest (US/Pop, FR/Pop, K-PoP, US-Politic, Adult); and *(iv)* Elites from a similar business sector in one or more countries (US/-Corp, US/TV, Global/Fun, Entertainment). *The cohesion of these elite communities suggests that there are social, cultural, or business forces that result in tighter connectivity among a group of elite users in these resilient communities.*

5.3 Elite Communities and their Temporal Evolution

Given the identified elite communities in the $S16$ and $S18$ snapshots, we next examine *how elite communities have evolved over time?* More specifically, we check whether and how an elite community in the $S16$ snapshot gets mapped to an elite community in the $S18$ snapshot. The change in the identity of the roughly 20% of top 10K elite users between the two snapshots coupled with any changes in their connectivity over time could affect the mapping of elite communities across these two snapshots. To this end, we count the number of overlapping nodes between a reference community $C_{16}(i)$ (in $S16$) and any elite community in 18 to detect the elite community $C_{18}(j)$ with the largest number of overlapping nodes (denoted by $overlap_{ij}$). We then map $C_{16}(i)$ to $C_{18}(j)$ with a mapping confidence computed as $overlap_{ij}/n_i$ where n_i denotes the number of $C_{16}(i)$ nodes that are still among the top 10K elites in $S18$. For each of the 14 elite communities in $S16$, Table 7 shows the community in $S18$ that the $S16$ community (labeled as "Ref. Comm.") is mapped to (labeled as "Mapped to") along with the number of overlapping nodes between them (labeled as #Nodes) and the computed mapping confidence in the last column. We observed two types of mappings between elite communities in the two snapshots. First, a majority of elite communities is mapped to new communities with a similar theme and moderate to high (70-99%) confidence, *e.g.*, IN, Arabic, Brazil, ID, and K-PoP. These are elite communities that have changed over time but their overall theme remains largely similar to how they appeared in $S16$. Second, elite communities whose themes have changed as they have significantly evolved or merged with other communities between the two snapshots (*e.g.*, TH, TR and US/TV). For the latter, the higher the mapping confidence, the larger the fraction of the original community that is collectively merged into a community with a new theme. While high mapping confidence (*e.g.*, TR) implies that a community completely merged into a new one, low mapping confidence (*e.g.*, TH) means that only a part of the community has merged. Note that a factor that affects this confidence is also the fraction of nodes in the reference elite communities that remain among the elite users in the new snapshot. *Overall, this analysis illustrate the relative stability of elite users and elite communities (along with their theme) over time.*

5.4 Elite Communities Across Different Views

We are also interested in knowing *whether and how an elite community's social cohesion and/or theme varies across different views within one and the same snapshot?* To answer this question, we consider 10 different views of the Twitter elite network (1K-ELITE, 2K-ELITE, ..., 10K-ELITE), identify the resilient communities in each view, and determine their social and location footprints. Furthermore, we keep track of the overlapping users between communities in consecutive views to establish their similarities. Leveraging a Sankey flow diagram, Figure 9 shows the relation-

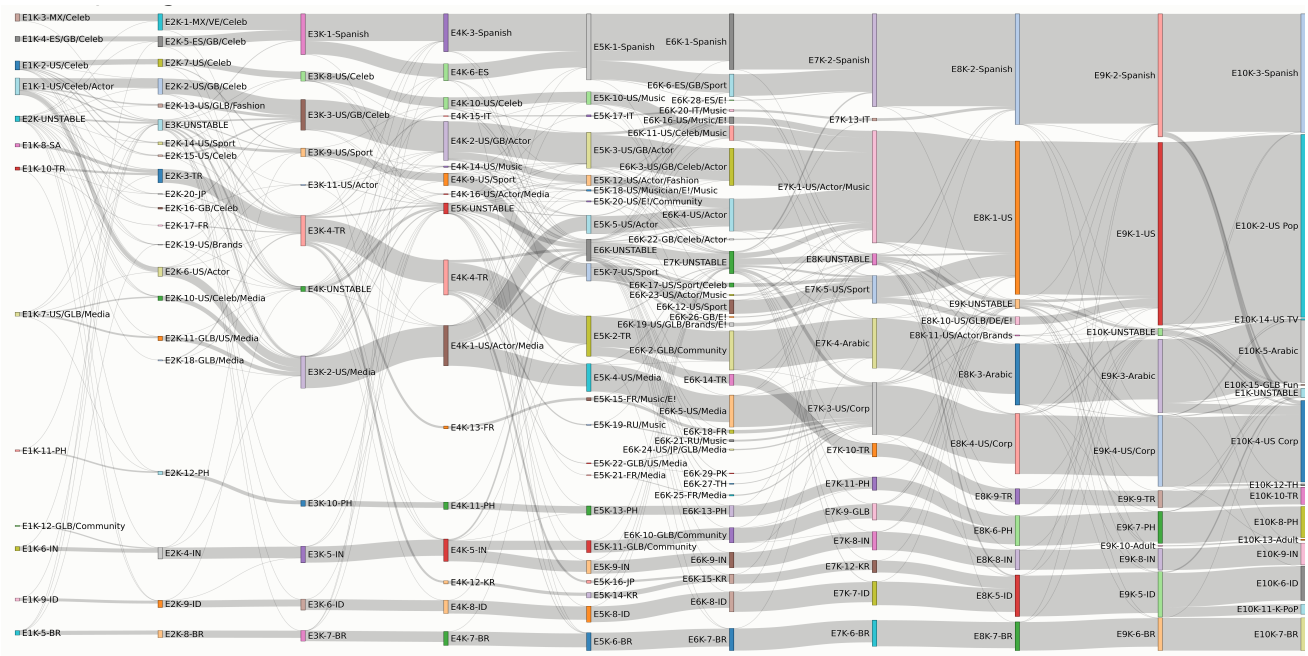


Fig. 9 The evolution of elite communities and their themes across different views of the elite network from 1K-ELITE through 10K-ELITE for the $S16$ snapshot of the elite network.

Ref. Comm.	Mapped to.	# Nodes(#)	Confidence (%)
US/Pop	US/Pop	1.5K	69.98
IN	IN	304	99.35
TH	K/Pop	11	45.83
PH	PH	172	62.32
GLB	US/Pop	2	50.0
US/TV	US/Corp	13	76.47
US/Corp	US/Corp	624	61.54
Arabic	Arabic	495	95.38
BR	BR	388	98.48
Unstable	Unstable	29	30.21
ID	ID	321	78.68
TR	Arabic	207	96.73
K/Pop	K/Pop	105	93.75
Adult	Adult	7	87.5
Spanish	Spanish	1057	70.75

Table 7 Mapping of individual elite communities in the snapshot $S16$ to an elite community (with the largest overlap) in the snapshot $S18$. The last column shows mapping confidence.

ships among elite communities in consecutive views as we expand the size of the elite network in the $S16$ snapshot from 1K to 10K accounts. The x-axis depicts the size of the elite network as it grows in steps of size 1K, and each group of vertically aligned bars represents the elite communities in a particular view. The length of each bar indicates the size of the corresponding community and its label shows the name of the community using the convention *view.size-theme*; *e.g.*, E9K-BR is a community in the 9K-ELITE network whose main theme is associated with Brazil. The gray horizontal strips between communities in consecutive views show the number of overlapping users (and thus similarity of themes) between those communities. Figure 9 illustrates that the collection of main themes among communities stabilizes as we consider larger views of the elite network in the $S16$ snapshot. We also observe that in general, elite communities exhibit strong social cohesion at all views. Figure 9 also shows that as new nodes are added to the network, many communities remain relatively stable (*e.g.*, E*K-*-BR, E*K-*-IN) while others merge (or split) across different views. While the former often have a consistent theme that may evolve over time (*e.g.*, "E6K-US/Media" evolves to "E7K-US/Corp" or "MX-Celeb" changes to "Spanish"), for the latter the theme often narrows (or broadens) as they split (or merge) (*e.g.*, "E9K-ID" splits into a larger "E10K-ID" and a smaller "E10K-KPop", and "E6K-Spanish" and "E6K-ES/GB/Sport" merge to form "E7K-Spanish"). In addition, the sizes of the elite communities increase as the elite network grows and their mapping across consecutive views becomes more pronounced (*i.e.*, the gray strips become wider and exhibit less splitting or merging between the largest three views).

To summarize, the relative stability of themes of elite communities across different views clearly indicates that these themes are not a side-effect of a particular network size but instead represent a robust social or geographic footprint of these communities. Furthermore, most of the elite communities from the 2016 snapshot are mapped with high confidence to elite communities with a similar theme in the 2018 snapshot. These findings confirm that elite communities with their specific themes represent "socially meaningful" and relatively easily-identifiable components of the Twitter elite network.

6 Community-Level Structure

The presence of pronounced communities with social cohesion in Twitter's elite network allows us to examine its connectivity structure at the community level. This coarse-grained view of the Twitter network is not only more manageable in terms of scale but also motivates a closer examination of the relationships among these communities. To this end, we explore in this section the notions of *(i)* direct friend-follower relationships, and *(ii)* indirect reachability. For each of these two notions of pairwise connectivity, we also compare and contrast the community level structure of the 10K-ELITE network between our two snapshots ($S16$ and $S18$) to gain insight into the evolution of the 10K-ELITE network at the community level and also examine the role of the set of unstable nodes.

6.1 Bias in Directed Pairwise Connectivity

A friend-follower relationship (*i.e.*, an edge) from user u to user v indicates that v is interested in following (and receiving tweets from) u. Similarly, the collection of such relationships from elite users in community C_i to their followers in community C_j captures the collective attention that C_i receives from C_j. Therefore, a directed connectivity structure among all elite communities reveals pronounced patterns of interest across these units. We emphasize that there are edges between all pairs of elite communities. Our goal is to examine *whether the pairwise connectivity between different pairs of elite communities exhibits any bias*. The heatmaps in Figure 10 and 11 illustrate the relative bias in *directed* connectivity between elite communities in the 10K-ELITE network for the $S16$ and $S18$ snapshots, respectively. More specifically, the color of cell (i,j) shows whether the number of directed edges from community C_i to community C_j is larger or smaller than the number of connections in a degree-preserving randomized version of the elite network[4]. Compared to the randomized structure, having more edges (shown in red) indicates a positive bias and having fewer edges (shown in blue) implies a negative bias. Given the directed nature of the connectivity, the resulting heatmap does not require to be symmetric. Also, note that the strip on the right side of each plot (indicating the scale for color coding) has a wider range for the $S16$ snapshot, *i.e.*, a particular color indicates a larger bias for $S16$ than $S18$. All communities are ordered based on their size from the bottom-up on the y-axis and from right to left on the x-axis.

These heat maps show a few interesting points about the nature of bias in the directed connectivity between elite communities in both snapshots and their evolution. For one,

[4] In a randomized degree-preserving version of the network, we randomly connect elite nodes while maintaining their in- and out-degrees.

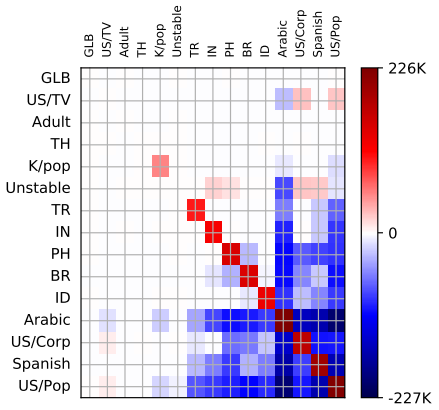


Fig. 10 Bias in directed connectivity between elite communities in 10K-ELITE for $S16$ snapshot

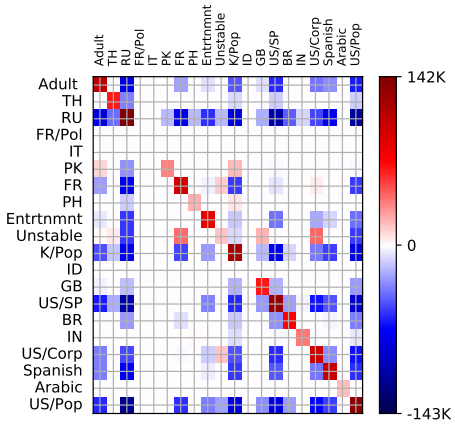


Fig. 11 Bias in directed connectivity between elite communities in 10K-ELITE for $S18$ snapshot

not surprisingly, all the largest positive values appear on diagonal position which indicates intra-community bias. However, there are also a few elite communities in both snapshots that do not exhibit any positive bias in their internal connectivity; *e.g.*, all the small communities (in the top-left portion of the heatmap) for $S16$ as well as the IT, FR/Pol, and ID communities for $S18$. This last group appears to have no negative or positive bias in their inter- or intra-community connections. Second, the larger the negative bias in the connectivity between the two communities is, the better separated they are. The heatmaps show that the largest four elite communities in snapshot $S16$ (*i.e.*, US/PoP, Spanish, US/-Corp, and Arabic) are clearly well separated from the rest of the elite communities. The heatmap associated with the $S18$ snapshot shows that except for US/Corp, the other three communities (*i.e.*, US/PoP, Spanish and Arabic) remain to be well separated in $S18$. A few other elite communities in $S18$ (*e.g.*, US/Sport and FR/Polit) exhibit a moderate level of separation. Third, interestingly the negative bias between elite communities (*i.e.*, blue pattern in the heatmap) for both snapshots is in general very symmetric, an indication that any separation between communities is largely reciprocal. Finally, observing some off-diagonal light-red cells shows

that there is a slight positive bias between the corresponding two communities (*e.g.*, US/TV and UC/Corp in $S16$ or K-PoP and FR in $S18$).

In summary, these heatmaps of bias in directed pairwise connectivity illustrate that the level of bias is symmetric between each pair, and some elite communities do not exhibit any bias in their internal or external connectivity while other elite communities are more clearly separated.

6.2 Indirect Pairwise Reachability

The "pairwise reachability" (*i.e.*, tight coupling) between two elite communities is an important aspect of connectivity that is not always correlated with the number of direct edges between them. To examine the notion of *pairwise reachability* between elite communities, we examine the outcome of the individual runs of the (Combo) community detection algorithm that we used to obtain the elite network. We recall that a detected community C_x in each run of Combo may include two (or more) resilient communities RC_i and RC_j. Such a "co-appearance" of RC_i and RC_j is an indication of their relative reachability or coupling. Therefore, the frequency of co-appearance for two resilient communities RC_i and RC_j in the communities identified by Combo (across 100 runs in Section 5) can be considered to be an informative measure for assessing their pairwise reachability.

Figures 12 and 13 summarize the pairwise reachability between all elite communities in 10K-ELITE where each circle represents a community for the $S16$ and $S18$ snapshots, respectively. The thickness of each undirected edge between a pair of nodes shows their pairwise reachability. We also label each edge with the corresponding frequency of co-appearance for nodes at both ends and only show edges whose weight is larger than 2 to keep the figure simple. In essence, these figures show the likelihood of bundling between all pairs of resilient communities in the outcome of each run of Combo.

These two figures reveal a few interesting points. First, a few elite communities (namely TR, BR, Arabic, Spanish) never co-appear with others in the $S16$ snapshot which reconfirms their clear separation from other elite communities. Two of these communities (Arabic, Spanish) remain largely separated in the $S18$ snapshot but the BR community is closely connected with a few new elite communities (IT, GB, FR) that appear in the second snapshot. Also, the US/PoP community has become more separated in the $S18$ snapshot. Second, there is a larger number of edges with a higher weight between communities in 2018 compared to 2016. This implies that more pairs of communities in the $S18$ snapshot co-appear with higher frequency despite the slightly lower density of its elite network (as we reported in Table 6). Third, we observe four distinct groups of elite communities that frequently co-appear together in the $S16$ snapshot, namely *(i)* US-Corp, US-TV and TH, *(ii)* ID and K-PoP, *(iii)* IN and PH, and *(iv)* US-Pop and GLB/Fun. The co-appearing elite communities in the $S18$ snapshot are quite different and consist of *(i)* ID and K/PoP, TH, RU, *(ii)* FR, IT, BR, GB, *(iii)* US/Corp, FR/Politics, *(iv)* Adult, US/Sport, Entertainment. While the first group has some similarity to the group in $S16$, all other groups either consist of new communities or new groupings of existing and new communities. In summary, despite the modest change in the identity of the top 10K elite nodes and edges between elites, the pairwise reachability between elite communities has significantly evolved between our two snapshots. The low frequency of co-appearance between two elite communities suggests that we can consider them as rather unrelated components of the elite network. Therefore, we can conclude that *the* 10K-ELITE *view of the Twitter elite network consists of roughly 9 separate components that are composed of groups and individual elite communities. But the composition of these components are very different between the two snapshots.*

6.3 The Role of Unstable Nodes

As we described in Section 5, unstable nodes do not consistently get grouped with any specific elite community since they have connections to users in many different clusters. This raises the question *whether unstable nodes serve as a bridge (or hub) between a pair/group of elite communities that result in their higher level of pairwise reachability?* To investigate this question, we consider all 155 (1.5%) unstable nodes in 10K-ELITE and determine their frequency of co-appearance with each elite community in the $S16$ snapshot. Figure 14 presents the frequency of co-appearance of these unstable nodes as a heatmap where the color of the cell (i, j) indicates the the frequency of co-appearance for unstable node j with elite community i[5]. A high frequency of co-appearance indicates tighter connectivity between a given unstable node and an elite community. Figure 14 reveals that four groups of these unstable nodes primarily co-appear with the following set of elite communities (from left to right): *(i)* US-Pop and Glb-Fun, *(ii)* US-Corp, TH, and US-TV, *(iii)* ID and K-PoP, *(iv)* PH and IN. These are exactly the same groups that showed significant pairwise reachability in Figure 12. This observation suggests that *these unstable nodes act as hubs and facilitate tighter coupling between the corresponding elite communities*. Figure 14 also shows two large groups of unstable nodes that are "hanging" off

[5] We use a simple reordering algorithm along the x-axis to group unstable nodes that have a similar co-appearance pattern. Note that the sum of the values in each column is not 100% since a co-appearance of an unstable node with multiple resilient communities is counted separately.

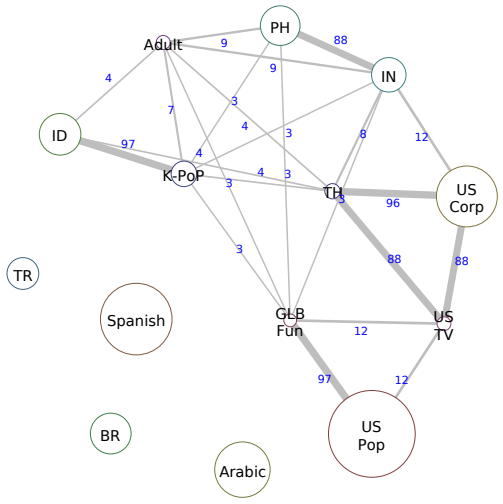


Fig. 12 The frequency of pairwise co-appearance of elite communities in $S16$ snapshot

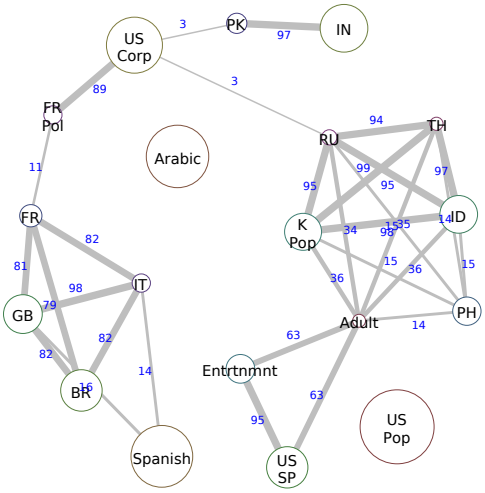


Fig. 13 The frequency of pairwise co-appearance of elite communities in $S18$ snapshot

(*i.e.*, primarily co-appear only with) the Spanish and Arabic elite communities and two smaller groups that primarily co-appear with the BR and TR communities.

7 Influence Among Elites

In this section, we investigate *how elite communities influence each other.* Prior studies on user influence have examined the influence of user u on all other (*i.e.*, mostly regular) users in a social network and have relied on metrics such as the total number of retweets, mentions, or replies by other users on posts originated by u. While these measures

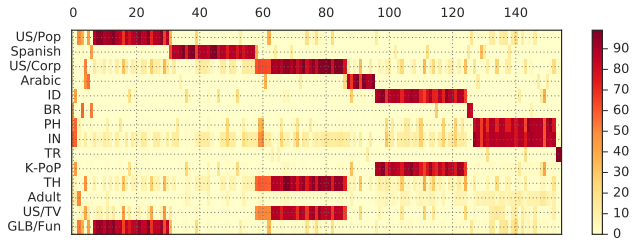


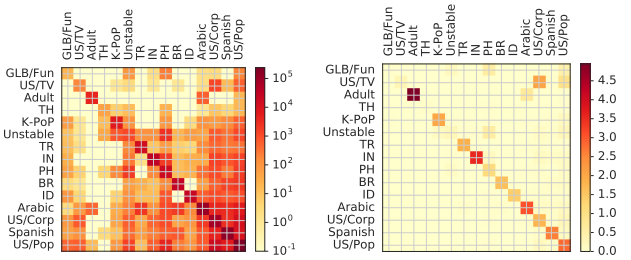
Fig. 14 Pattern of co-appearance for the 150 unstable nodes (on the x-axis) with the 14 elite communities (on the y-axis) across the 10K-ELITE network in snapshot $S16$

of user engagement and user degree are generally correlated [9], the ranking of influential users based on user engagement and user connectivity measures (*e.g.*, PageRank) are not strongly correlated [12,7].

There are four important differences between our analysis of influence among elite communities and prior studies [7,12,3,8]. First, we only focus on influence between elite users (rather than all users) in a network. Second, to quantify pairwise influence between elite users, we consider a modified version of an engagement-based metric based on *retweets* and *replies*. Third, we characterize cross-influence at the granularity of elite communities as well as individual users. Lastly, we examine the relationship between community-level influence and community level importance in the elite network.

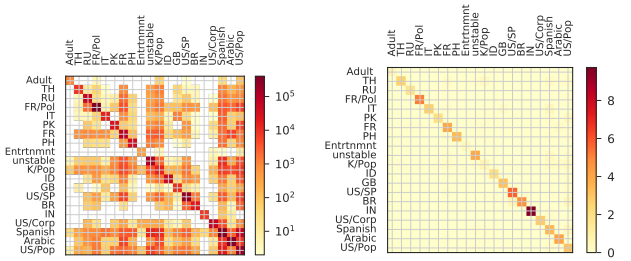
Most prior engagement-based influence measures for user u use the total number of retweets or replies by all other users to u's post (*e.g.*, [7]). We capture the overall influence of an elite user u (in terms of retweet or reply) on all other elite users with two metrics, namely *(i) number of influenced elites*: the number of unique elite users who have retweeted (or replied to) at least one of u's original tweets (an indication of how widespread u's influence is); and *(ii) aggregate influence*: the summation of the fractions of any other elite user's captured tweets that are retweets of (or replies to) tweets originally generated by u (an indication of the aggregate magnitude of u's influence for this user). More specifically, this aggregate influence for user u is defined as $AggUserInfl(u) = \sum_{v \in Elite} RT_{u \to v}/N_v$ where $RT_{u \to v}$ denotes the number of times that user v retweeted (or replied to) user u and N_v is the total number of v's tweets. We can also define the retweet (or reply) influence of community C_i on community C_j as the summation of all pairwise influences for any user in C_i on any user in C_j; that is, $AggCommInfl(C_i, C_j) = \sum_{v \in C_i} \sum_{w \in C_j} RT_{v \to w}/N_w$

To conduct this analysis, we collect all available tweets of all accounts in the 10K-ELITE network for both the $S16$ and $S18$ snapshots. Our dataset for snapshot $S16$ ($S18$) contains more than 31M (29M) tweets where 6.5M (6M) of them are retweets and 5M (4.6M) are replies.



(a) Num. of Retweets (b) Retweet Influence (%)

Fig. 15 Directed pairwise community-level influence in the $S16$ snapshot

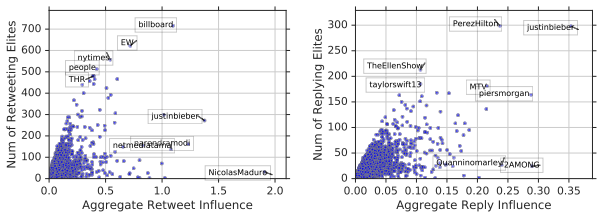


(a) Num. of Retweets (b) Retweet Influence (%)

Fig. 16 Directed pairwise community-level influence in the $S18$ snapshot

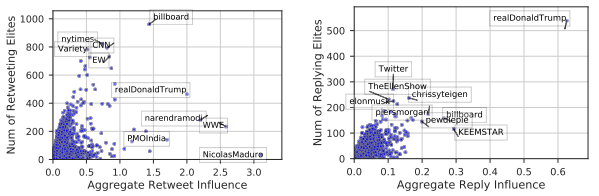
7.1 Community-Level Influence

In Figures 15(a) and 16(a), the color of cell (i,j) indicates the absolute number of times that a user in elite community i has retweeted tweets originated by users in elite community j. These heatmaps show that in both snapshots, the members of each elite community primarily influence other members of their own community. Furthermore, larger communities also influence other (smaller) communities. Interestingly, the level of influence is generally symmetric between different pairs of elite communities. Since these absolute metrics could be biased towards larger communities, Figures 15(b) and 16(b) present the normalized views of influence where the color of the cell (i,j) indicates the percentage of tweets by users in community i that is a retweet of tweets originated by users in elite community j. This normalized view provides a more proper representation of the influence between elite communities. Surprisingly, this measure has non-zero values mostly on the diagonal cells. The only noticeable exception to this dominant pattern is the retweet influence of US-Corp on US-TV in snapshot $S16$. Furthermore, the level of influence within elite communities is not a function of their size. In particular, elite users in the Adult, IN and Arabic communities in the $S16$ snapshot and elite users in the IN, US/Sport and Fr/Politics in the snapshot $S18$ have the most retweet influence on other elites in the same community. We observe very similar result for reply influence (available in [15]) with elites in K-PoP, IN, and TR showing the most reply influence on their community members.



(a) Retweet Influence (b) Reply Influence

Fig. 17 Influence of individual elite users on all other elites in $S16$



(a) Retweet Influence (b) Reply Influence

Fig. 18 Influence of individual elite users on all other elites in $S18$

To summarize, these results demonstrate that both the retweet and reply influence of elite users are primarily contained within their own elite community. Moreover, this feature of influence among elites does not change over time.

7.2 User-Level Influence

To gain a more micro-level understanding, we also characterize the patterns of pairwise user-level influence among elites. Figures 17 and 18 show for both snapshots the influence for individual elite users in a scatter plot where each point presents a user, its x-value indicates the user's aggregate retweet (or reply) influence and its y-value shows the number of unique elite users influenced by that user. When comparing the two figures, we observe some characteristics that are common between the two snapshots. On one end of the spectrum, we observe some elite users (*e.g.*, @nicolasmaduro, President of Venezuela) that exhibit large retweet influence but on only a small number of elite users. On the other end of the spectrum, some elite accounts, associated with news (*e.g.*, @Billboard, @EW and @nytimes) exhibit a lower aggregate influence but on a large number of elite users. Furthermore, there are few accounts that show high retweet influence with respect to both dimension (*e.g.*, @justinbieber in snapshot $S16$ and @realDonaldTrump in snapshot $S18$).

Figures 17(b) and 18(b) also show a number of characteristics of the reply influence in both snapshots. For one, both dimensions of the reply influence exhibit a lower value than the corresponding value for the retweet influence. Moreover, in both snapshots, we notice that the most influential replying elite users are in general celebrities in the entertainment industry (*e.g.*, @PerezHilton, @justinbieber @MTV

while the most influential retweeting elites are often news agencies and political figures. Finally, when comparing each influence measure for both snapshots, we observe that the range of influence for both dimensions of each measure has slightly increased over time. We also see the emergence of @realDonaldTrump (the US president) as a very influential elite user by both measures between the first and second snapshot. While his increased retweet influence seems plausible and justified, the large increase in his reply influence that is significantly larger than all other celebrities and entertainment accounts comes as a surprise.

*In summary, this analysis illustrates that both dimensions of retweet or reply influence are equally important to gain meaningful insight into the nature of the influence for individual elite accounts. While the general patterns of user-level influence among elites does not change over time, the influence of certain accounts (*e.g.*, @realDonaldTrump) has dramatically increased with respect to both measures between our two snapshots.*

7.3 Influence vs. Importance of Elite Communities

To get a broader view of influence for individual elite communities, we examine the influence of their nodes based on different metrics. An intriguing question is *whether the relative influence of elite users in a community is related to (or affected by) their relative position in the elite network?* We address this question by comparing the rank of a group of users in each elite community based on their retweet, reply and PageRank [6] measure, respectively. While the retweet and reply measures represent different aspects of user influence, the PageRank metrics capture the relative importance (*i.e.*, centrality) of a user location in the elite network. Figures 19 and 20 present the summary distributions of user rank based on retweet (purple) and reply (blue) influence as well as PageRank (green) for individual elite communities (including unstable nodes) in the 10K-ELITE network for the $S16$ and$S18$ snapshot, respectively.

For each elite community, we examine whether the summary distribution of PageRank across its users has overlap with the summary distribution of the rank based on retweet and reply influence. Such an overlap for each influence measure indicates that the users' position affects their level of influence. Using this recipe for examining Figures 19 and 20, two observations stand out. First, for most elite communities in both snapshots, the rank of users based on their location in the graph is comparable with the rank based on their retweet and reply influence, which suggests that there is a close relationship between the location and the influence of users. However, there are a few communities where their level of influence diverges from their location in the elite network. For example, the Arabic and Adult communities exhibit higher reply influence and the IN community

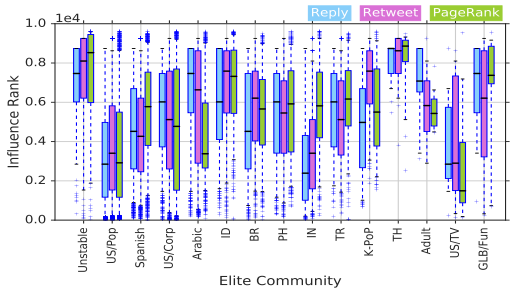


Fig. 19 Distribution of the rank of users in each of the elite communities in the 10K-ELITE network for the $S16$ snapshot based on two measures of influence and PageRank.

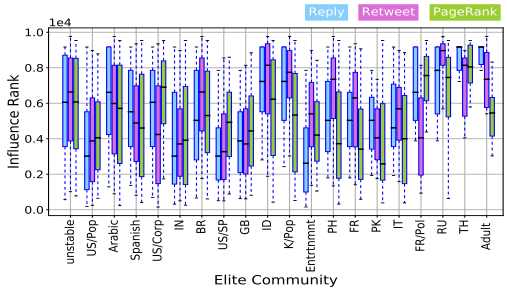


Fig. 20 Distribution of the rank of users in each of the elite communities in the 10K-ELITE network for the $S18$ snapshot based on two measures of influence and PageRank.

has lower reply influence compare to their locations in the elite network for the $S16$ snapshot. Similarly, the FR/Politics community exhibits lower and the RU (Adult) community shows higher retweet (reply) influence than their locations in the network for the $S18$ snapshot. Second, contrasting the location/influence characteristics of users in common communities between both snapshots shows more alignment between location and influence measures across different communities in 2018 compared to 2016. For example, the different measures have become more aligned for the Arabic and IN communities in the $S18$ snapshot. The only noticeable exception to this trend is the Adult community where the reply and retweet influence of its users has increased over time and has become more misaligned with their location between the two snapshots.

In summary, the observed patterns basically reflect the nature of the overall influence of an individual elite community and its association with its location along with any changes in these features over time.

8 From Communities to Partitions

One of the main benefits of identifying elite communities is to use them to gain more insight into the structure of the Twitter network as a whole. As reported earlier (see Section 2), more than 80% of regular (*i.e.*, non-elite) Twitter users follow at least one elite user in the 10K-ELITE network (*i.e.*, have at least one elite friend). This high visibility of elite users coupled with our observation of the existence

of socially-cohesive elite communities raises the question *whether the regular users can be broadly divided into meaningful partitions where each partition contains regular users associated with a particular elite community?*

For an initial investigation of the association between regular users and elite communities (see [15] for more details), we randomly select 10K regular users as representative samples of all Twitter users. Referring to the fraction of elite friends of a regular user u that are located in elite community c as u's *belonging factor* with respect to community c, we observe that a significant fraction of regular users has more than 70% of their elite friends in a single elite community. Therefore, these regular users can be reliably mapped to the elite community that has the largest belonging factor. Finally, to test our hypothesis that the collection of regular users that are mapped to a single elite community can be viewed as a *"shadow partition"* of that elite community, we consider 100K randomly selected friend-follower relationships between regular users and then map the regular users at both ends to their corresponding elite communities. We observe that 35.2% of these relationships are between users in different shadow partitions. This ratio is very similar to the fraction of relationships between elite users that are located in different elite communities. In short, we argue that since elite communities can be used in practice to partition regular users, they are capable of providing a meaningful macroscopic view of the entire Twitter OSN.

9 Conclusion & Outlook

In this paper, we present a socially-informed characterization of the Twitter elite network that contains the top-10K most-followed accounts and their friend-follower relationships. We devise a new technique for efficiently capturing Twitter's elite network, and after annotating each node in the resulting network with its social attributes, we identify resilient elite communities. We show that these elite communities exhibit social cohesion with a clear theme and represent therefore socially meaningful entities of the Twitter OSN. We then characterize both the connectivity and influence among elite communities and show that grouping regular users based on their association with an elite community result in "shadow partitions" whose inter-connectivity mirrors that of the corresponding elite communities. As a result, these shadow partitions can be viewed as extensions of the elite communities across the entire Twitter graph, *i.e.*, the elite communities represent a socially-meaningful coarse-grained view of both the Twitter elite network and the Twitter network as a whole.

Our future plans to extend this work include an in-depth study of the temporal evolution of the Twitter network at the level of elite communities (including their associated social themes). We also plan to extend the notion of shadow partitions by leveraging individual elite communities as landmarks and cluster regular users based on their level of connectivity to all elite communities.

Acknowledgements We would like to thank Hooman Mostafavi for his help in collecting the second snapshot of the Twitter elite network. This material is based upon work supported by the National Science Foundation under grant IIS-0917381 and CNS-1320977.

References

1. M. Al-Garadi. Analysis of Online Social Network Connections for Identification of Influential Users: Survey and Open Research Issues. *ACM Computing Surveys*, 51, 2018.
2. K. Avrachenkov, N. Litvak, L. O. Prokhorenkova, and E. Suyargulova. Quick detection of high-degree entities in large directed networks. In *Proc. of ICDM*. IEEE, 2014.
3. E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts. Everyone's an influencer: quantifying influence on twitter. In *Proc. of WSDM*. ACM, 2011.
4. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), 2008.
5. B. Bollobás. *Modern graph theory*, volume 184. Springer Science & Business Media, 2013.
6. S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *Proc. of WWW*. ACM, 1998.
7. M. Cha, H. Haddadi, F. Benevenuto, and P. K. Gummadi. Measuring user influence in twitter: The million follower fallacy. *ICWSM*, 2010.
8. M. Cha, A. Mislove, and K. P. Gummadi. A measurement-driven analysis of information propagation in the flickr social network. In *Proc. of WWW*. ACM, 2009.
9. D. Easley and J. Kleinberg. *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press, 2010.
10. S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3), 2010.
11. R. Gonzalez, R. Cuevas, R. Motamedi, R. Rejaie, and A. Cuevas. Assessing the evolution of google+ in its first two years. *IEEE/ACM Transactions on Networking (ToN)*, 24(3):1813–1826, 2016.
12. H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a Social Network or a News Media? In *Proc. of WWW*. ACM, 2010.
13. J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney. Community Structure in Large Networks: Natural Cluster Sizes and The Absence of Large Well-Defined Clusters. *Internet Mathematics*, 6(1), 2009.
14. R. Motamedi, R. Rejaie, D. Lowd, W. Willinger, and R. Gonzalez. Inferring coarse views of connectivity in very large graphs. In *Proc. of COSN*, 2014.
15. R. Motamedi, S. Rezayi, R. Rejaie, R. Light, and W. Willinger. Characterizing twitter elite communities. Tech. Report CIS-2016-15, www.cs.uoregon.edu/Reports/TR-2016-015.pdf, Univ. of Oregon, 2016.
16. R. Motamedi, S. Rezayi, R. Rejaie, and W. Willinger. On characterizing the twitter elite network. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 234–241, 2018.
17. T. Puranik and L. Narayanan. Community detection in evolving networks. In *Proc. of ASONAM*, 2017.
18. R. Rejaie, M. Torkjazi, M. Valafar, and W. Willinger. Sizing up online social networks. *IEEE network*, 24(5):32–37, 2010.

19. M. Rosvall and C. Bergstrom. Maps of information flow reveal community structure in complex networks. In *Proc. of the National Academy of Sciences*, 2007.
20. S. Sobolevsky, R. Campari, A. Belyi, and C. Ratti. General optimization technique for high-quality community detection in complex networks. *Physical Review E*, 90(1), 2014.
21. D. Stutzbach, R. Rejaie, N. Duffield, S. Sen, and W. Willinger. On unbiased sampling for unstructured peer-to-peer networks. *IEEE/ACM Transactions on Networking*, 17(2), 2009.
22. M. Torkjazi, R. Rejaie, and W. Willinger. Hot today, gone tomorrow: on the migration of myspace users. In *Proceedings of the ACM workshop on Online social networks*, pages 43–48, 2009.
23. M. Valafar, R. Rejaie, and W. Willinger. Beyond friendship graphs: a study of user interactions in flickr. In *Proceedings of the ACM workshop on Online social networks*, pages 25–30, 2009.
24. Wikipedia. Sankey diagram. `https://en.wikipedia.org/w/index.php?title=Sankey_diagram&oldid=740785912`, 2019. Accessed: 2019-04-24.
25. J. Yang and J. Leskovec. Overlapping community detection at scale: a nonnegative matrix factorization approach. In *Proc. of WSDM*, 2013.