DREXEL UNIVERSITY
College of
Computing & Informatics

Investigating Coordinated 'Social' Targeting of High-Profile Twitter Accounts

Hunter Scott Heidenreich Munif Ishad Mujib Jake Ryland Williams

College of Computing and Informatics, Drexel University

Motivation

Data collection was initiated to observe possible effects of the 2020 US Presidential Democratic Primary debates on Twitter follower counts of candidates. The Twitter Rest API was utilized to construct a streaming acquisition of follower count changes for 97 prominent Twitter accounts including those of the presidential candidates. However, rather than post-debate effects, initial exploratory analysis of the data revealed two types of intriguing patterns.

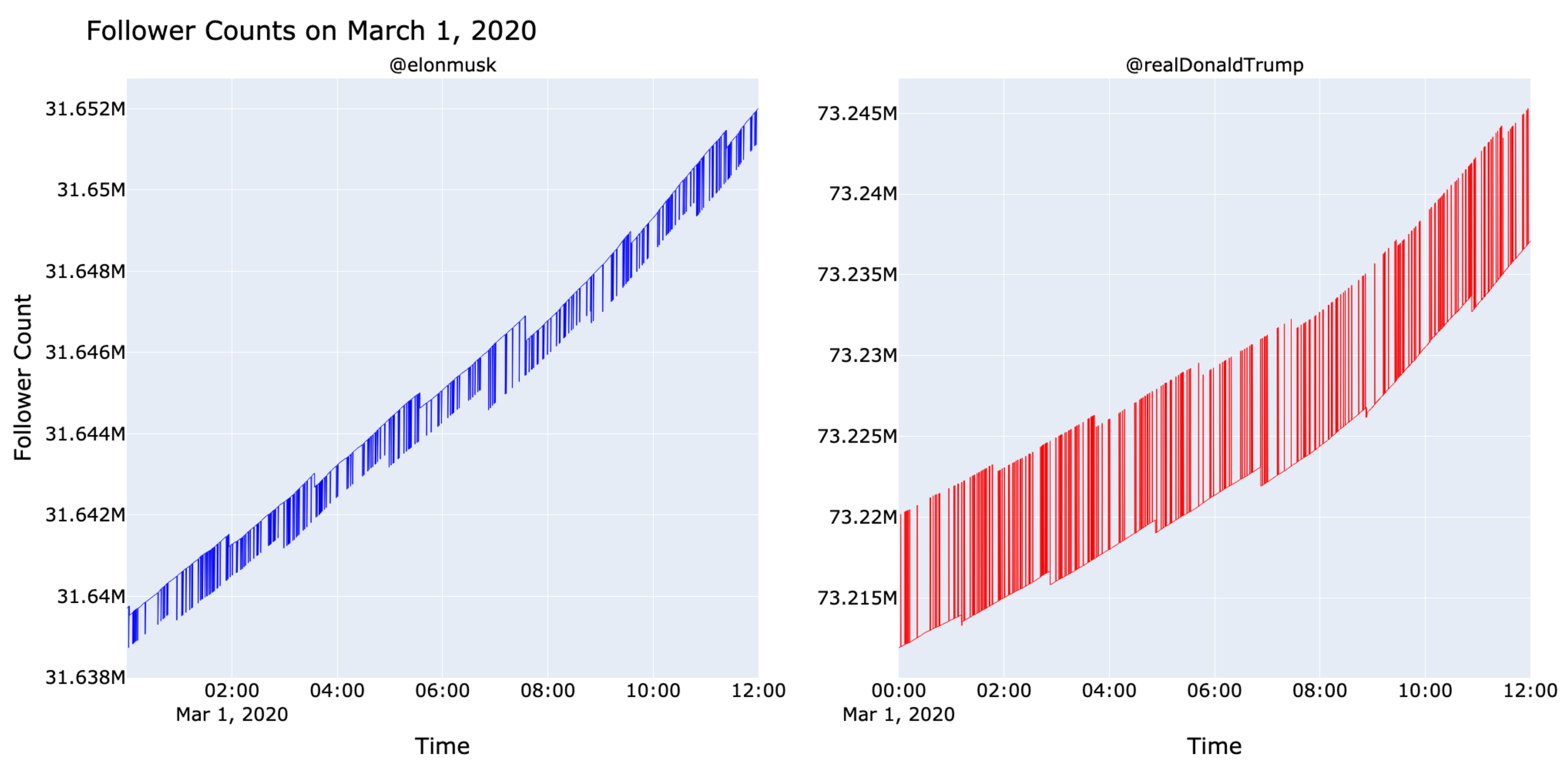

Figure 1: Follower counts of Elon Musk and Donald Trump demonstrating spike and sawtooth patterns.

- **Spikes**: Sharp follower count increases and decreases that immediately disappear, i.e. the follower count returns to the pre-spike level.
- **Sawteeth**: Sharp follower count decreases that do not return the follower count to previous levels, i.e. a more permanent change compared to spikes.

Repeated occurrences of both of these patterns strongly suggest automated behavior, since organic follower growth or loss is unlikely to occur in such a discrete manner. Table 1 displays a summary of follower count data collected between 2/6 and 5/23 from the top 5 accounts in 3 categories.

Table 1: Observed Spike and Sawtooth Characteristics

Username	Followers (millions)	Spikes/Day	Avg. Spike Effect	Net Spike Effect	Sawteeth/Day	Avg. Sawtooth Effect	Net Sawtooth Effect
Presidential Candidates							
realDonaldTrump	75.73	608	7,422	481,740,537	25	-7,907	-22,915,819
BernieSanders	11.41	411	98	3,044,831	147	-361	-5,699,337
JoeBiden	4.70	361	-70	-2,511,603	48	-123	-650,322
amyklobuchar	1.01	108	-38	-392,359	12	-28	-67,696
GovBillWeld	0.09	46	-18	-43,303	11	-113	-479,920
Individuals							
elonmusk	32.63	574	-924	-58,088,500	32	-1,053	-3,205,884
Cristiano	83.51	527	-1,156	-63,668,866	25	-1,460	-3,229,595
ArianaGrande	72.29	526	-456	-26,673,829	39	-1,619	-4,826,954
narendramodi	54.63	506	-4,679	-95,641,313	26	-5,434	-8,564,685
justinbieber	110.71	492	-144	-8,839,477	47	-2,032	-7,782,180
Organizations							
NASA	35.91	505	-387	-22,288,451	38	-806	-4,123,770
WhiteHouse	21.12	502	-253	-13,426,879	25	-512	-1,385,918
BBCBreaking	42.37	493	-564	-29,563,485	21	-861	-1,871,289
CNN	46.43	459	-1,037	-16,622,269	42	-1,217	-2,341,954
nytimes	45.83	449	-13	1,183,160	39	-746	-3,399,156

Operating under the assumption that the majority of effects from these patterns are inorganic, one or more automated network or networks of Twitter accounts in the 10^6 scale, at least, must exist. These groups of accounts must also be exhibiting a specific type of behavior: repeatedly following and unfollowing high-profile accounts. These **circulating accounts** are key to understanding the phenomena of spikes and sawteeth.

Uncovering Circulating Accounts

A cycling download of the most recent 10,000 followers for each tracked account was initiated and maintained. API rate limitations resulted in an approximately 3-hour period between downloads for each user. For many accounts, the number of circulating followers found were in the 10^6 range (Table 2). Critically, when follower growth rates are so high that significantly more followers than 10,000 are acquired within the 3-hour window, this method fails to capture the most recent, short-period circulation activity. The much lower-than-expected circulation numbers for @realDonaldTrump are potentially a result of this limitation.

Table 2: Observed Circulation Characteristics

Username	Circulating Followers	Circulations	Circulations/Day
Presidential Candidates			
realDonaldTrump	574,617	588,517	5,350
BernieSanders	1,725,767	4,938,692	44,897
JoeBiden	1,393,163	4,673,670	42,487
amyklobuchar	239,093	3,917,608	35,614
GovBillWeld	13,496	2,358,950	21,641
Individuals			
elonmusk	1,962,410	2,732,585	24,841
Cristiano	2,460,784	3,905,542	35,504
ArianaGrande	1,757,980	1,968,667	17,896
narendramodi	1,381,224	2,653,892	24,126
justinbieber	2,223,687	2,855,490	25,959
Organizations			
NASA	2,263,533	4,423,285	40,211
WhiteHouse	2,118,902	3,966,168	36,056
BBCBreaking	2,218,925	4,469,857	40,635
CNN	2,136,211	3,233,881	29,398
nytimes	2,087,851	4,673,142	42,483

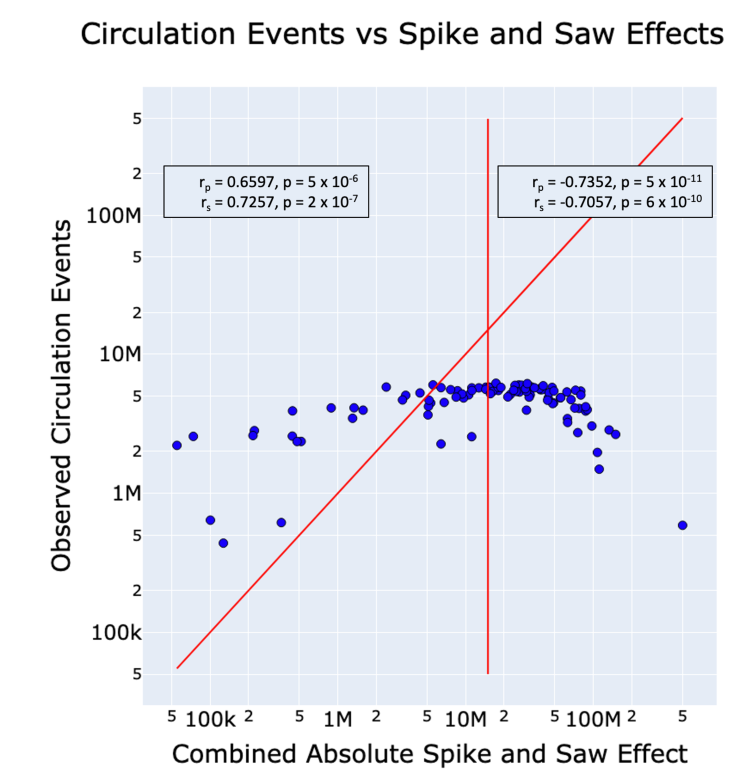

Figure 2: As the total spike and saw effect increases beyond a critical point near 15,000,000, more circulation effects presumably escape detection. 39 low-effect accounts with positive correlation and 58 high-effect accounts with negative correlation are seen.

Figure 2 plots observed circulation events against observed spike and saw effects, demonstrating through correlation a seemingly (concave down) functional relationship between the two. Thus, these two entirely-separately measured effects are related. The non-monotonic behavior exhibited here can potentially be interpreted as follows: for accounts with high spike and saw effects, capturing all circulations becomes a challenge (because of rapid burying rates). This is evidenced by the under-estimation which gets 'worse' as a falling trend in figure 2. At the other end, non-zero veracity in the ordering of follower lists (by follow time) 'bakes in' some over-estimation for all, but is only observable for small spike-saw effect accounts who generally receive too-few new followers to drown out this relatively marginal (by orders of magnitude) over-estimation. @realDonaldTrump is the extreme outlier at bottom right.

Figure 3 illustrates observed circulation activity on the final two Democratic candidates' accounts. Despite having half the followers of @BernieSanders, circulation activity on @JoeBiden often matched or exceeded @BernieSanders'.

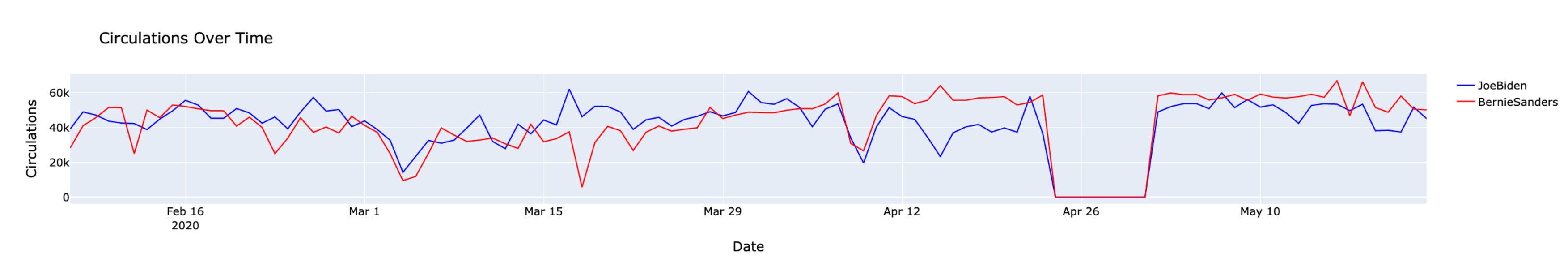

Figure 3: Circulation activity on @JoeBiden and @BernieSanders. Gap between 4/26-4/30 is due to data loss.

Inspecting follower lists revealed a further unexpected phenomenon: the abnormal abundance of "ancient" Twitter accounts among recent followers, especially among recent followers of @realDonaldTrump.

Uncovering Ancient Accounts

We define an **"ancient" account** as any with 8-digit or smaller user ID (i.e. accounts created on or before 12/09).

Characterization and Scale

Since inception, we have observed **∼2 million** ancient accounts. This is *at least* 2% of "ancient" Twitter and ∼6.5% of the DAU in 2010. We observe similar shifts in "attention" of these accounts in this sub-population, like a marked shift from Bernie Sanders to Joe Biden in early March, exhibited in Figure 4.

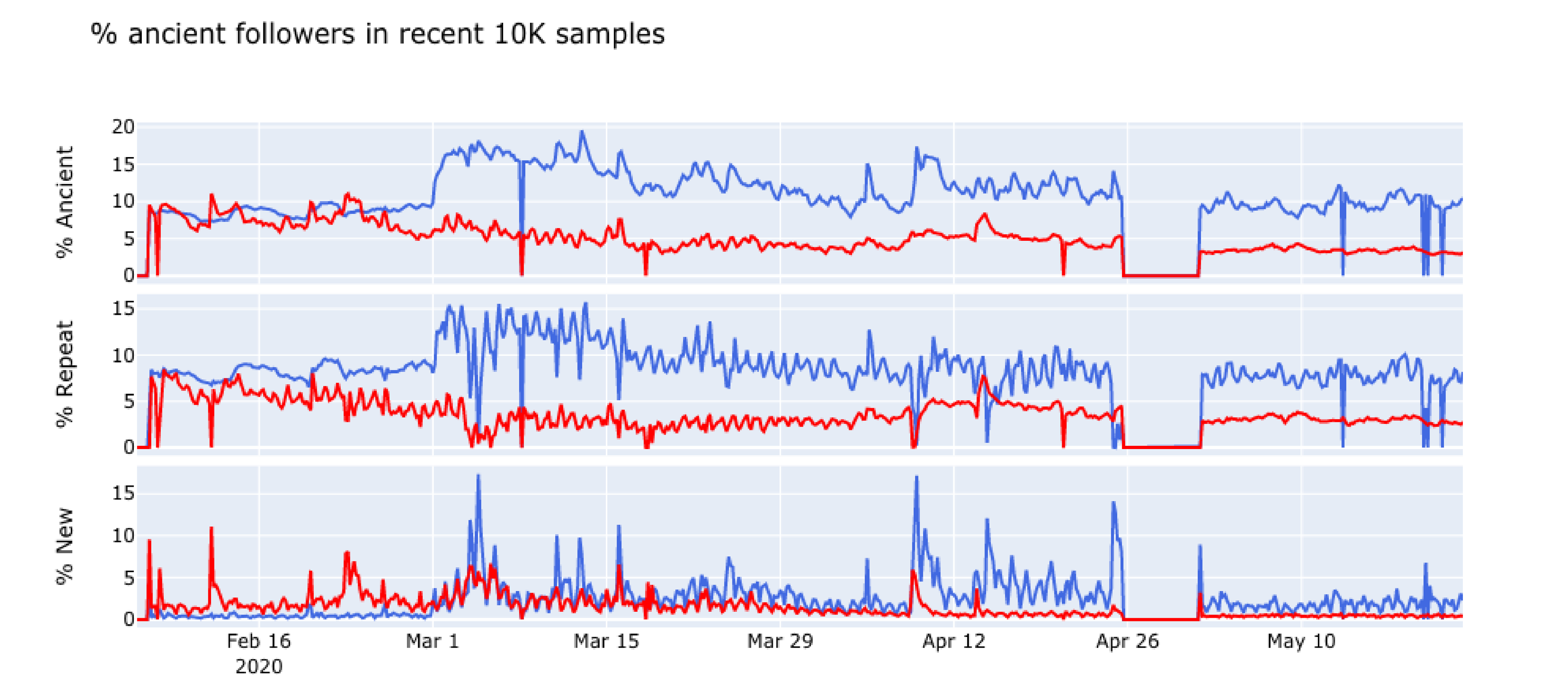

Figure 4: % of ancient followers observed in top 10K samples over time. Note the seeming shift from Sanders (red) to Biden (blue) at the start of March. Gap between 4/26-4/30 is due to data loss.

Large Time Gaps

Of the ancient accounts observed, we monitor the timelines of all with 7 or fewer digits and all 8-digit accounts following @realDonaldTrump. Of those with tweet gaps ≥ a year, many have only just "re-awoken" as seen in Fig. 5.

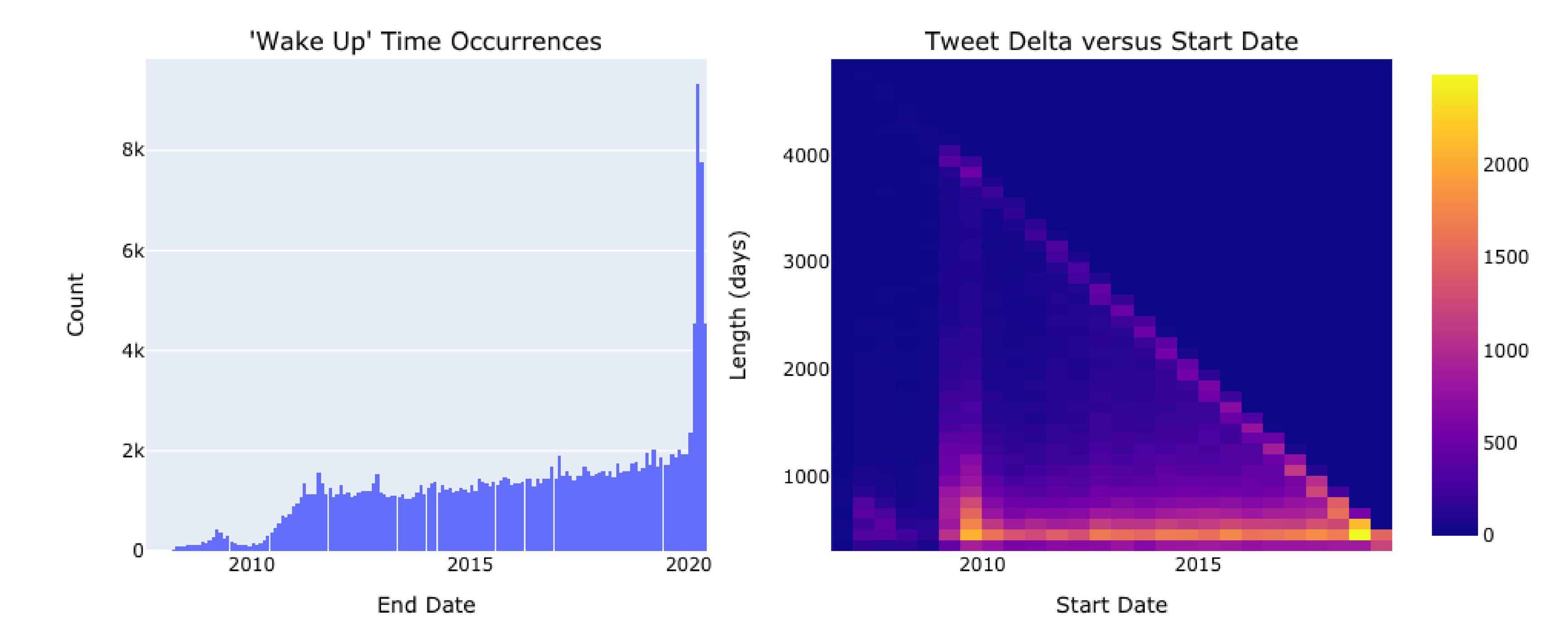

Figure 5: (Left) Histogram of end dates of tweet deltas that are longer than a year. A large number have recently started tweeting again, aligning with early US lockdown procedures and reaching a peak during 3/21-3/23. (Right) Regardless of start date, there is a strong linear correlation between length of gap and its end date due to the effects of many accounts beginning new tweet behaviors.

What's Next?

Here, we highlight some of the directions we intend to focus future work on.

- **Network Linkages** Through lists, friends, and followers, networks can be constructed from these accounts and could help to resolve questions about coordination and connectivity.
- **Intent** If we've observed a botnet, it's important to establish its intent. Through at-scale semantic analysis of profiles and tweets, clustering of social behaviors, and network connections, we hope to better understand intent and coordinated activities.
- **Mechanism** If spikes and sawteeth are a controlled mechanisms, what are their intended effects, e.g., are these intended to sway public perception of a user, or somehow affect Twitter's platform (e.g., through search or recommendation)?

Acknowledgements

This material is based upon work supported by the National Science Foundation under grant no. 1850014.