

Lorem ipsum lorem ipsum lorem ipsum lorem ipsum lorem
 ipsum lorem ipsum lorem ipsum lorem ipsum lorem ipsum
 lorem ipsum lorem ipsum lorem ipsum lorem ipsum lorem
 ipsum lorem ipsum lorem ipsum lorem ipsum lorem ipsum
 lorem ipsum lorem ipsum lorem ipsum lorem ipsum lorem
 ipsum lorem ipsum lorem ipsum lorem ipsum lorem ipsum
 lorem ipsum lorem ipsum lorem ipsum lorem ipsum lorem
 ipsum lorem ipsum lorem ipsum

Lorem ipsum lorem ipsum lorem ipsum lorem ipsum lorem
 ipsum lorem ipsum lorem ipsum lorem ipsum lorem ipsum
 lorem ipsum lorem ipsum lorem ipsum lorem ipsum lorem
 ipsum lorem ipsum lorem ipsum lorem ipsum lorem ipsum
 lorem ipsum lorem ipsum lorem ipsum lorem ipsum lorem
 ipsum lorem ipsum lorem ipsum lorem ipsum lorem ipsum
 ipsum lorem ipsum lorem ipsum.

We examined the possible ways that bots can engage with

human users on the Twitter platform, determining that these vectors consisted primarily of:

- Mentioning a user in a tweet
- Retweeting a user
- Using a popular hashtag or phrase
- Following a user
- Favoriting another user's tweet

We collected data about how many accounts each account was following, how many accounts followed them, and how many tweets each account had favorited. This user data, averaged across a cluster, gave insight into how each cluster of users tended to use the Twitter platform.

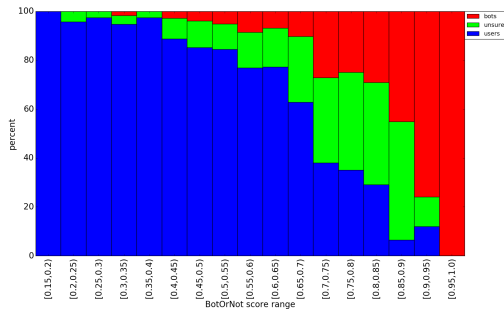
In addition to retrieving this user data, we also obtained the latest tweets made by each account, with a total of 625,053 tweets. We could then determine whether the tweet had been retweeted from another user, contained links, mentioned other users, and which hashtags were used.

Using these social behaviors, we were able to determine how each cluster and category of user tended to interact with other users. Focusing bot detection on these interactions could result in improved efficiency for spam removal services or other bot-related studies.

4. RESULTS & DISCUSSION

Examining the correctness of *BotOrNot* scoring was the first part of the analysis performed. We manually determined whether a number of accounts were bots, humans, or indeterminate. For the majority of accounts with a *BotOrNot* score over 50%, either we were unable to conclusively determine if the account was automated, or we determined that the account was definitely a bot. In addition, accounts with a *BotOrNot* score less than 50% were almost entirely found to be human users, as seen in Figure 1.

Figure 1: Manual identification of accounts



We next examined how an account's overall *BotOrNot* score aligned with the average of its corresponding category subscores, confirming that, with some deviations, the *BotOrNot* score was a better predictor of whether or not a user's account was automated. From Figure 2, we see that classifying the accounts based on the overall *BotOrNot* score is more effective than using the averaged category subscores.

Our unsupervised machine learning algorithms were able to similarly separate bot users and human users, as seen in

Figure 2: *BotOrNot* Scores vs Category Subscores

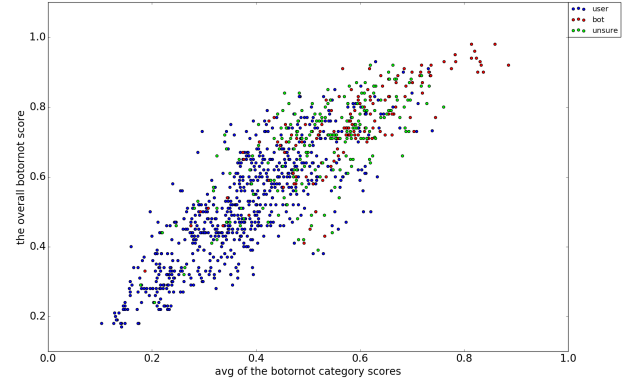
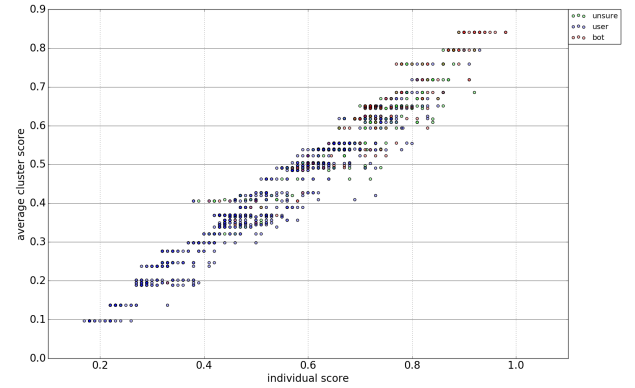


Figure 3. Clusters are visible on the same horizontal line, and these clusters tend to be either mainly bots or mainly humans.

Figure 3: *BotOrNot* Clustering vs Average Scores



4.1 Mentions

As seen in Figure 4, we found that Twitter bots tended to not mention specific users, with most bot clusters having fewer than 17 mentions per user. Twitter's guidelines for bots are particularly explicit about this aspect of the service:

If your application creates or facilitates automated reply messages or mentions to many users, the recipients must request or otherwise indicate an intent to be contacted in advance.[9]

In Figure 5, we examine the amount that accounts in clusters that mention users actually do so. If bots tended towards a similar rate of mentioning users as humans did, we would expect to see the data remain similar across clusters from Figure 4 to Figure 5. However, since bots tend to mention at a higher frequency when they do mention other users, clusters containing bots have comparatively higher mentioning rates in Figure 5. In clusters of humans, the mentioning

Category	[0, 20%)	[20%, 40%)	[40%, 60%)	[60%, 80%)	[80%, 100%]
Human	10	150	277	204	21
Bot	0	1	17	70	53
Indeterminate	0	5	37	114	41

Table 1: Number of manually classified accounts within *BotOrNot* score ranges

characteristics of those users is more uniform. In contract, in the bot clusters, users tend to have either a medium amount of mentions, or none at all.

Figure 4: Mentions Per Cluster

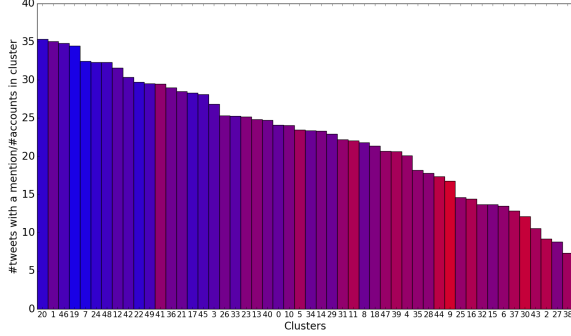
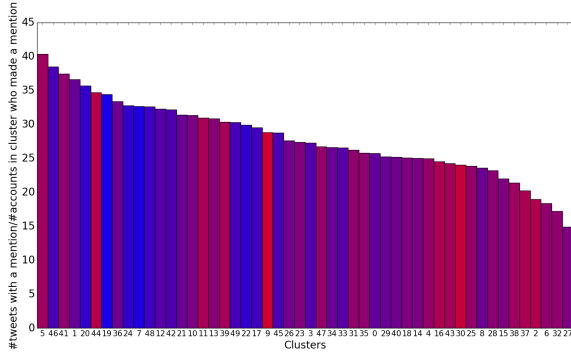


Figure 5: Mentions Per User



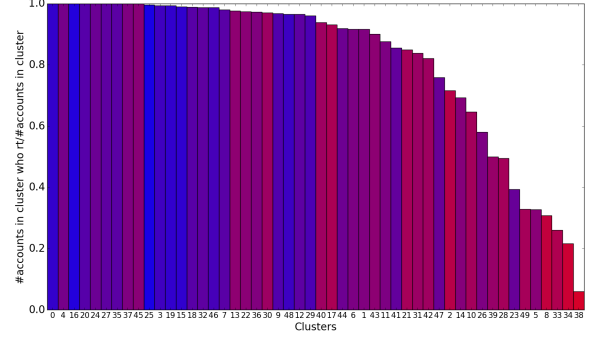
4.2 Retweets

Similar to automated mentions, Twitter’s guidelines explicitly forbid automated retweeting:

Automation of Retweets often leads to spam and other negative user experiences; therefore, Retweeting in a bulk or automated manner is prohibited.[9]

Very few users classified as bots mention other users or retweet their tweets, which indicates that Twitter is closely monitoring these methods of user interaction; either very few bots are being created to automate these actions or they are rapidly banned from the service.

Figure 6: RTs Per Cluster



4.3 Hashtags

Due to some selection bias for the Twitter accounts we scanned, a disproportionate number of users tweeted using hashtags related to a shared interest, such as independent game development.

However, examining how many times each user in a cluster used a hashtag, as shown in Figure 7, gave a more indicative view of hashtags used for spamming, such as “fifa15coins” and a number of sexually explicit hashtags. Many of these spammed hashtags were only tweeted by a single user, which indicates that the creators of these spam accounts attempt to avoid overlap in which hashtags they are spamming. While Twitter’s terms of service forbids automatically posting into the trending topics, it does not forbid automation using hashtags, allowing for these bots to find commonly popular hashtags and spam them without much apparent risk.

A large number of users tweeted with the hashtag of “Finances,” but under further inspection a majority of these accounts were verified, indicating that while these accounts may exhibit spamming behavior, they were likely controlled by humans.

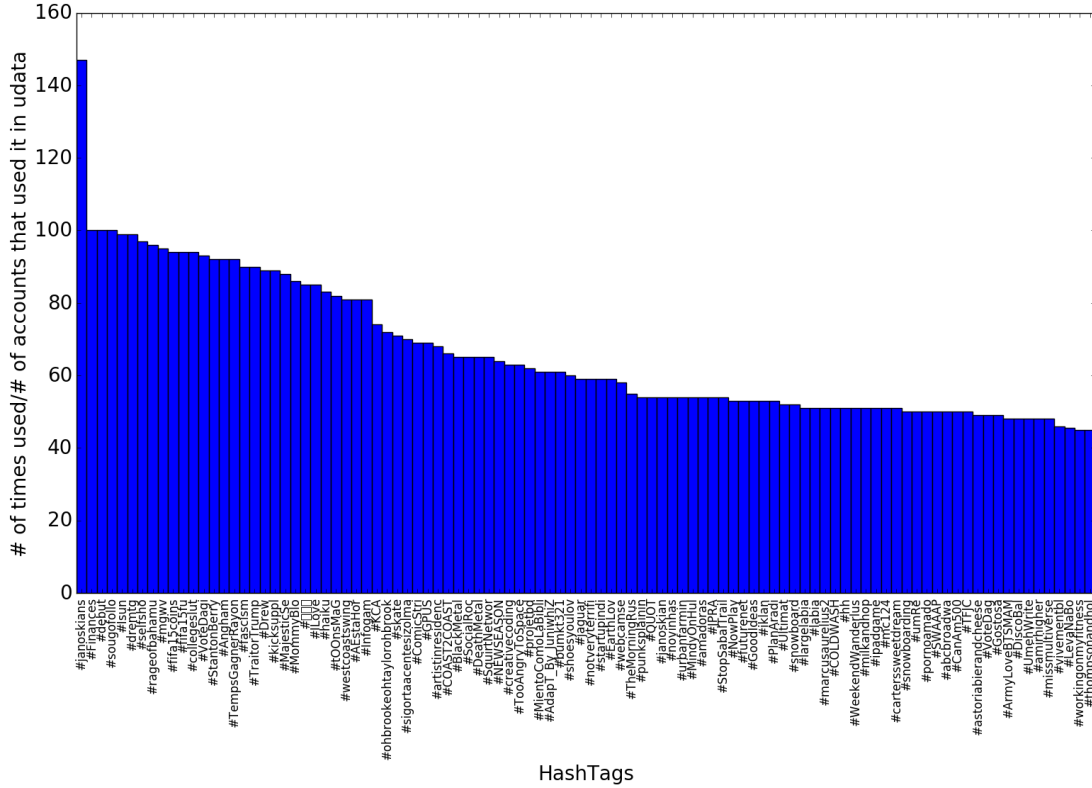
Examining a cluster comprised mainly of bots, such as cluster 33 in Figure 11, confirms that these automated users are far more likely to repeatedly use the same sexually explicit hashtags, in addition to very general hashtags such as “NBA.”

4.4 Following

While Twitter forbids automated following, it’s one of the most common methods bots use to gain attention from human users. Most bots are following over 10,000 users; they also have similarly exaggerated numbers of followers.

We noticed one cluster of users sharing an extremely large number of both followers and accounts they followed; upon further inspection, the majority of these accounts were promotional accounts. Since they tend to follow back, these ac-

Figure 7: Hashtags Per User



counts become a prime target for sockpuppet accounts that want to appear legitimate; both the following and follower tabs are filled with users who have the default Twitter profile picture and no tweets, with many of their creation dates within the last month.

4.5 Favoriting

Examining how often users favorited tweets gave an interesting trend; bots tended to either favorite very few tweets, or a very large number of them, with human users hovering in between the two extremes.

We manually examined the users in clusters 22 and 33. Cluster 22 had the largest number of favorites, while cluster 33 favorited a large number of tweets and scored rather heavily in the *BotOrNot* assessment. Cluster 22 was comprised mainly of businesses and other Twitter “personalities” such as YouTubers. These users likely either search their names and favorite tweets including that text, or favorite the tweets that mention them as a way of quickly responding to fans.

Cluster 33, however, was comprised of accounts that tweeted links to related sites with little commentary besides relevant hashtags. These “aggregator” Twitter accounts search for hashtags and tweets relating to the topic that they post and then favorite those tweets. This behavior may pass under Twitter’s radar because the accounts select phrases to search for that are relatively uncommon, simply reflecting the intermittent behavior of other users.

5. CONCLUSION

Our study focused on analyzing interaction between human and bot Twitter users. We used *BotOrNot*, combined with unsupervised machine learning, to cluster users and determine how bots gain visibility with their target audience. We determined that bots are generally unlikely to engage with human users beyond simply following them and using hashtags. However, when bots *do* interact with users, they do so without moderation, resulting in bots tending towards behavioral extremes.

5.1 Further Work

Although we were able to manually identify advertising links, when we retrieved data on individual Tweets we did not expand Twitter’s shortened URL format. This made media, such as photos, appear identical to other links, since Twitter represents media as URLs. In addition, the sample size for manual classification was small by necessity; setting up a web service such as Mechanical Turk to crowdsource this account classification would improve analysis and clustering.

Furthermore, while several accounts were discovered that exhibited bot-like behavior such as spamming a hashtag, a number of these accounts were verified by Twitter, especially those run by so-called “financial consultants.” While a closer examination of these users was outside the scope of this project, they seem closely related to the issue of spam-bots, and further discussion as to whether these accounts violate Twitter’s terms of service seems warranted.

While we sought to provide an analytical overview of our sample users, a deeper statistical analysis of individual clus-

Figure 8: Hashtags Per User (Cluster 33)

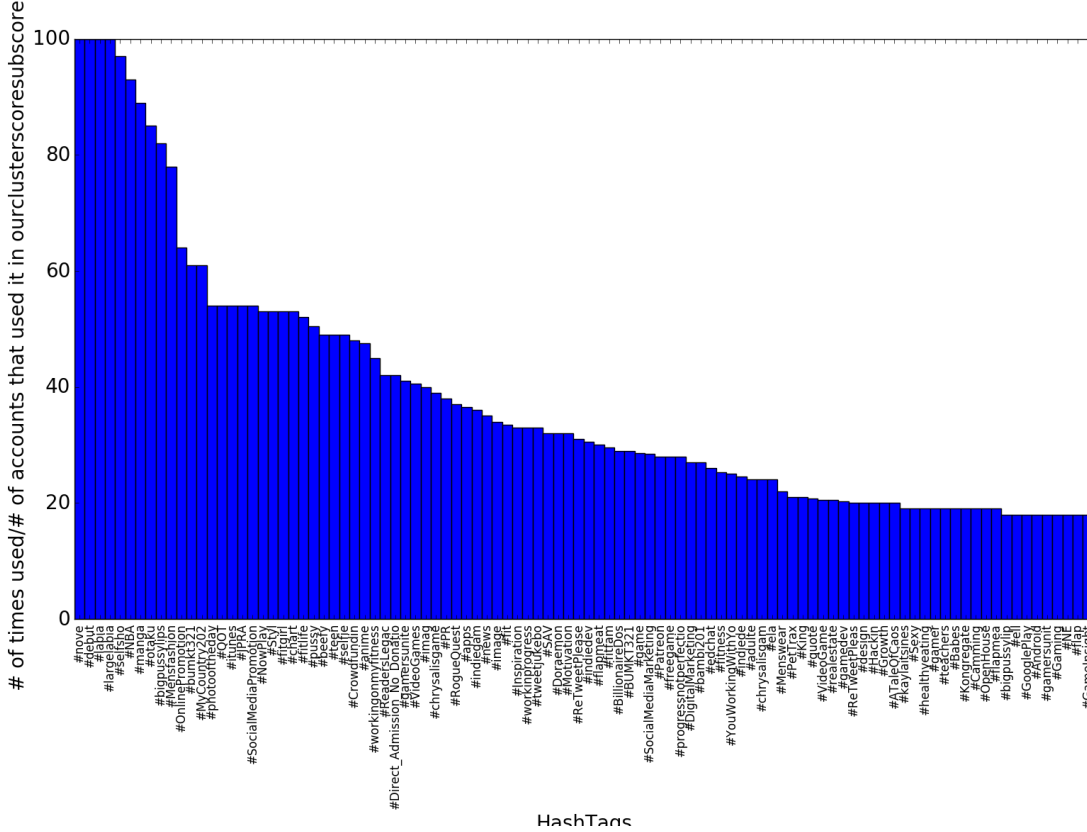


Figure 9: Following Per Cluster

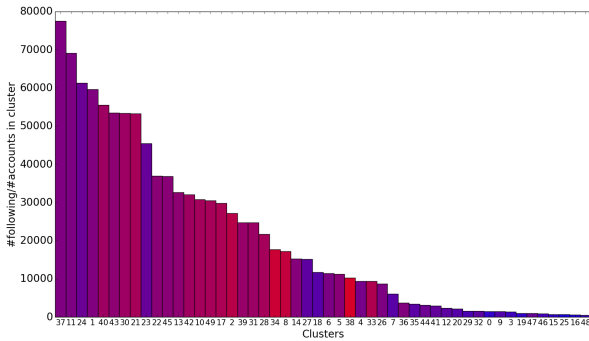
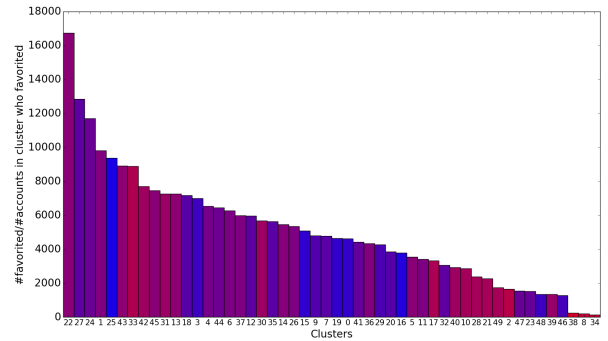


Figure 10: Favorites Per Cluster



ters and bots is warranted, based on several of the characteristics we identified. In particular, analyzing the distribution of mentions across individual clusters of bots and clusters of humans would lead to greater insight into that aspect of their behavior.

Finally, Twitter allows for users to quote other Tweets, as a commentary layered upon a retweet. This is unlikely to be a vector by which bots interact with humans, but it would provide a further view of Twitter users' behavioral patterns and its absence could possibly indicate automated behaviors.

6. REFERENCES

- [1] S. Aral and D. Walker. Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science*, 57(9):1623–1639, 2011.
- [2] A. Bessi and E. Ferrara. Social bots distort the 2016 U.S. Presidential election online discussion. *First Monday*, 21(11), 2016.
- [3] N. Bilton. Social media bots offer phony friends and real profit, nov 2014.
- [4] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia.

Detecting automation of twitter accounts: Are you a human, bot, or cyborg? *IEEE Trans. Dependable Secur. Comput.*, 9(6):811–824, Nov. 2012.

- [5] C. A. Davis, O. Varol, E. Ferrara, A. Flammini, and F. Menczer. Botornot: A system to evaluate social bots. *CoRR*, abs/1602.00975, 2016.
- [6] J. P. Dickerson, V. Kagan, and V. S. Subrahmanian. Using sentiment to detect bots on twitter: Are humans more opinionated than bots? In *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*, pages 620–627, Aug 2014.
- [7] J. Ratkiewicz, M. Conover, M. Meiss, B. Goncalves, A. Flammini, and F. Menczer. Detecting and tracking political abuse in social media, 2011.
- [8] G. Stringhini, C. Kruegel, and G. Vigna. Detecting spammers on social networks. In *Proceedings of the 26th Annual Computer Security Applications Conference, ACSAC '10*, pages 1–9, New York, NY, USA, 2010. ACM.
- [9] Twitter. Automation rules and best practices, apr 2016.
- [10] C. Xiao, D. M. Freeman, and T. Hwa. Detecting clusters of fake accounts in online social networks. In *Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security, AISec '15*, pages 91–101, New York, NY, USA, 2015. ACM.

APPENDIX

A. CONTRIBUTIONS

Alic Szecei provided data retrieval methods for Twitter accounts, programmed the unsupervised machine learning, and wrote the data analysis.

Willem DeJong programmed BotOrNot score retrieval, retrieved data for Twitter accounts to store in SQL databases, and created many of the graphs and charts.

B. MISC. DATA

Figure 11: Manual identification of accounts

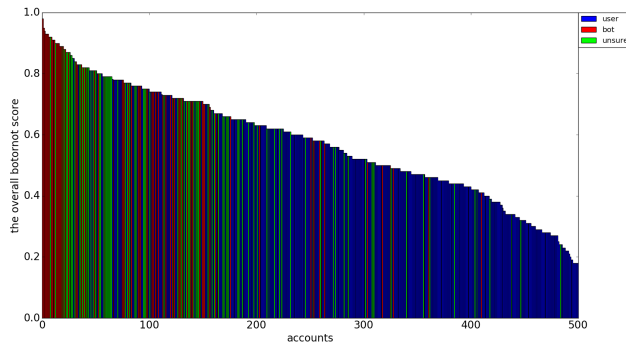


Figure 12: Followers Per Cluster

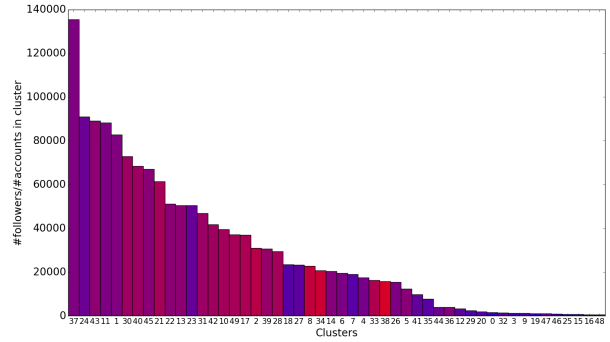


Figure 13: Times Listed Per Cluster

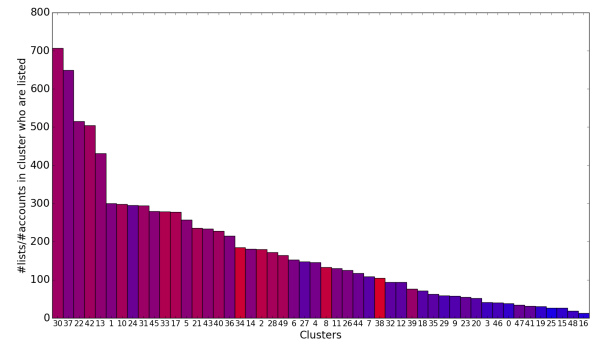


Figure 14: Presence In Lists Per Cluster

