

# Of Bots and Humans (on Twitter)

Zafar Gilani\*, Reza Farahbakhsh†, Gareth Tyson‡, Liang Wang\*, Jon Crowcroft\*

\*Computer Laboratory, University of Cambridge, Cambridge, UK, {szuhg2, lw525, jac22}@cam.ac.uk

†CNRS Lab UMR5157, Institut Mines Telecom, Paris, France, reza.farahbakhsh@it-sudparis.eu

‡Queen Mary University of London, London, UK, g.tyson@qmul.ac.uk

**Abstract**—Recent research has shown a substantial active presence of bots in online social networks (OSNs). In this paper we utilise our past work on studying bots (*Stweeler*) to comparatively analyse the usage and impact of bots and humans on Twitter, one of the largest OSNs in the world. We collect a large-scale Twitter dataset and define various metrics based on tweet metadata. Using a human annotation task we assign ‘bot’ and ‘human’ ground-truth labels to the dataset, and compare the annotations against an online bot detection tool for evaluation. We then ask a series of questions to discern important behavioural bot and human characteristics using metrics within and among four popularity groups. From the comparative analysis we draw differences and interesting similarities between the two entities, thus paving the way for reliable classification of bots, automated political infiltration and advertisement campaigns.

**Index Terms**—information dissemination; social network analysis; bot characterisation; behavioural analysis

## I. INTRODUCTION

Bots (automated agents) exist in vast quantities in online social networks (OSNs). They are created for a number of purposes, such as news, emergency communication, marketing, link farming,<sup>1</sup> political infiltration,<sup>2</sup> spamming and spreading malicious content. According to independent research, 51.8% of all Web traffic is generated by bots.<sup>3</sup> Similarly, OSNs have seen a massive surge in their bot population; Twitter reported<sup>4</sup> in 2014 that 13.5 million accounts (5% of the total user population back then) were either fake or spam. The rise of bots on Twitter is further evidenced by a number of studies [12], [15], [9], [6], and articles discussing bots.<sup>5</sup>

This constitutes a radial shift in the nature of content production, which has traditionally been the realm of human creativity (or at least intervention). Although there have been past studies on bots (§II), we are particularly interested in exploring their role in the wider social ecosystem, and how their behavioural characteristics differ from humans. This is driven by many factors. The limited cognitive ability of bots clearly plays a major role, however, it is also driven by their diverse range of purposes, ranging from curating news to answering customer queries. This raises a number of interesting questions regarding how these bots operate, interact and affect online content production: What are the typical behaviours of humans and bots, in terms of their own

activities as well as the reactions of others to them? What interactions between humans and bots occur? How do bots affect the overall social activities? The understanding of these questions can have deep implications in many fields such as social media analysis, systems engineering, etc

To answer these questions, we have performed a large-scale measurement and analysis campaign on Twitter (§III).<sup>6</sup> We focus on bots in Twitter because it largely exposes public content, and past studies indicate a substantial presence of bots [5]. Addressing existing problems with automated bot detection algorithms, we utilise a human annotation task to manually identify bots, providing us with a large ground-truth for statistical analyses. We analyse the most descriptive features from the dataset, as outlined in a social capitalist study [8], including six which have not been used in the past to study bots. Through our comprehensive approach, we offer a new and fundamental understanding of the characteristics of bots vs. humans, observing a number of clear differences (§IV). For example, we find that humans generate far more novel content, while bots rely more on retweeting. We also observe less intuitive trends, such as the propensity of bots to tweet more URLs, and upload bulkier media (e.g. images). We also see divergent trends between different popularity groups (based on follower counts), with, for example, popular celebrities utilising bot-like tools to manage their fanbase.

We analyse the social interconnectedness of bots and humans to characterise how they influence the wider Twittersphere. We observe that, although human contributions are generally considered more important via typical metrics (e.g. number of likes, retweets), bots still sustain significant influence. Our experiments confirm that the removal of bots from Twitter could have serious ramifications for information dissemination and content production on the social network. Specifically, we simulate content dissemination to find that bots are involved in 54.59% of all information flows (defined as the transfer of information from one user to another user). As well as providing a powerful underpinning for future bot detection methods, our work makes contributions to the wider field of social content automation. Such understanding is critical for future studies of social media, which are often skewed by the presence of bots.

<sup>1</sup>Link farming – <http://bit.ly/2cXhfBv>

<sup>2</sup>Bots distort U.S. presidential election – <http://bit.ly/2l3VzGf>

<sup>3</sup>Bot traffic report 2016 – <http://bit.ly/2kzZ6Nn>

<sup>4</sup>Twitter’s 2014 Q2 SEC – <http://bit.ly/1kBx4M8>

<sup>5</sup>Bots in press and blogs – <http://bit.ly/2dBAIbB>

<sup>6</sup>Datasets from this study will be made available upon notification.

## II. RELATED WORK

Two main streams of research are relevant to our work: (*i*) social, demographical and behavioural analyses of either bots or humans; and (*ii*) the impact of bots in limited social environments. Bot detection *per se* is not the focus of this paper. Many such techniques are focussed on discerning anomalous from normal, spam from non-spam, and fake from original, but they fail to distinguish (or compare) the types of users. Note that a user can be a human and still be a spammer, and an account can be operated by a bot and still be benign.

**Social analysis of bots or humans.** Most related to our work are behavioural studies of bots or humans. For example, [10] studied the *infiltration strategies* of social bots on Twitter using a manual approach. They use three metrics to quantify the infiltration of social bots: followers, popularity score, and message-based interaction (other users favouriting, retweeting, replying or mentioning the bot). They found that bots can successfully evade Twitter defences (only 38 out of their 120 bots got suspended over the course of 30 days). A similar work [3] studied users' behaviours in response to bot infiltration, calculating an infiltration success rate of 80%. They concluded that a successful infiltration can result in privacy breaches (to expose more than publicly available data), and that the security defences of Facebook are ineffective in detecting infiltration.

Unlike these works, we do not aim to monitor the success of bot infiltration. Rather we are interested in understanding the behavioural differences of bots and humans. That said, there is work that has inspected bot or human behaviour in isolation. For example, [4] examined the retweet behaviour of people, focussing on *how people tweet*, as well as *why and what people retweet*. The authors found that participants retweet using different styles, and for diverse reasons (*e.g.* for others or for social action). This is relevant to our own work, as we also study retweets. In contrast, we directly compare retweet patterns of bots and humans (rather than just humans).

Thus, our work provides further insights on important differences and striking similarities between bots and humans in terms of *account lifetime*, *content creation*, *content popularity*, *entity interaction*, *content consumption*, *account reciprocity*, and *content dissemination*. To the best of our knowledge, we are the first to perform this methodical comparison of key metrics across these two types of Twitter accounts.

**Social influence of bots.** The above studies primarily inspect the characteristics of bots. There has also been work inspecting the social influence of bots, *i.e.* how other users react to them. In [2], the authors use a bot on aNobii, a social networking site aimed at readers, to explore the *trust*, *popularity* and *influence* of bots. They show that gaining popularity does not require individualistic user features or actions, but rather simple social probing (*i.e.* bots following and sending messages to users randomly). The authors also found that an account can circumvent trust if it is popular (since popularity translates into influence). Closely related is [14], which develops models to identify users who are *susceptible* to social bots, *i.e.* likely to follow and interact

with bots. The authors use a dataset from the Social Bot Challenge 2011, and make a number of interesting findings, *e.g.* that users who employ more negation words have a higher susceptibility level. We take inspiration from this work and extend exploration to the Twitter platform. However, instead of infiltrating a social network with "honeypot" bot(s), we study the characteristics of existing bots. We argue that this provides far broader vantage into real bot activities. Hence, unlike studies that focus on the influence of individual bots (*e.g.* the Syrian Civil War [1]), we gain perspective on the wider spectrum of how bots and humans operate, and interact.

## III. METHODOLOGY

We use and build upon our past work *Stweeler*<sup>7</sup> [11] for data collection, pre-processing, human annotation, and analysis. We define a 'bot' as any account that *consistently* involves automation over the observed period, *e.g.* use of the Twitter API or other third party tools, performing actions such as automated likes, tweets, retweets, *etc*. Note that a *tweet* is an original status and not a retweet, a *retweet* is a tweet which has 'RT' in text, and a *status* is either a tweet or a retweet. Also note that *content* on Twitter is limited to whatever is contained within a tweet: text, URL, image, and video.

### A. Data Collection

We selected Twitter because it is open, large-scale and is known to contain a wide breath of bot activities. We collect data on bot and human behaviour for 30 days in April 2016 from the Streaming API. Since we collect all the tweets within a time period  $T$ , we are certain we have comprehensive insights within  $T$ . This resulted in approximately 65 million tweets, with approximately 2 to 2.5 million recorded per day. We then extracted the accounts and all associated metadata (*e.g.* account age) from tweets. In total, we recorded information on 2.9 million unique accounts. In this study, in addition to known metrics (age, tweets, retweets, favourites, replies and mentions, URL count, follower-friend ratio, *etc*) we also analyse a set of six novel metrics not explored in past bot research. These include: *likes per tweet*, *retweets per tweet*, *user replies and mentions*, *activity source count*, *type of activity sources*, and *size of content uploaded*. The selection of features is driven by [8] and, to our knowledge, this is the most comprehensive study to date.

### B. Data Pre-Processing

Our data contains a range of accounts in terms of their popularity (*i.e.* number of followers). During preliminary analysis we found that the purpose and activity of an account differs based on its popularity. Hence, we partition profiles into four popularity groups to enable a deeper understanding. These are as follows:

- **G<sub>10M+</sub> – celebrity status:** This is the subset of Twitter users with the highest number of followers, *i.e.* >9M followers. These are the most popular users that hold

<sup>7</sup>*Stweeler*– <https://github.com/zafargilani/stcs>

celebrity status and are globally renowned. Popular and credible organisations (*e.g.* CNN, NetGeo) use these accounts for various purposes, which makes them free of spam, thus having high credibility and trustworthiness.

- **$G_{1M+}$  – very popular:** This subset of Twitter users is amongst the most popular on the platform, *i.e.* 900K to 1.1M followers. These users are close to celebrity status and global recognition (*e.g.* nytfood, pcgamer).
- **$G_{100k}$  – mid-level recognition:** This subset represents popular accounts with mid-level recognition (*e.g.* CBSPhilly, DomusWeb), *i.e.* 90k to 110k followers.
- **$G_{1k}$  – lower popularity:** This subset represents more ordinary users, *i.e.* 0.9k to 1.1k followers. These users (*e.g.* hope\_bot, Taiwan\_Agent) form a large base and, though they show lower individual and accumulated activity, they do form the all-important tail of the distribution.

Our dataset is a representative sample of Twitter users, where each metric follows Gaussian distribution.  $G_{10M+}$  and  $G_{1M}$  are similar in their characteristics (*cf.* §IV) and constitute 0.65% of the total 105k accounts we partitioned in the dataset.  $G_{1k}$  represents the bulk of Twitter, constituting 94.40% of the total partitioned accounts.  $G_{100k}$  bridges the gap between the most popular and least popular groups, constituting 4.93% of the total partitioned accounts. A possible  $G_{10k}$  would be similar to  $G_{1k}$ , and a possible  $G_{50k}$  will be similar to  $G_{100k}$ .

### C. Bot Classification

To compare bots with humans, it is next necessary to identify which accounts are operated by bots. We initially experimented with BotOrNot [7], a state-of-the-art bot detection tool (to the best of our knowledge, this is the only available online tool). However, inspection of the results indicated quite high levels of inaccuracy. Hence, we chose to take a manual approach instead — we made this design choice to have a smaller but more reliable set of classifications. We employed human participants to perform a *human annotation task*<sup>8</sup> to identify bots and humans.

We recruited four undergraduate students for the purposes of annotation, who classified the accounts over the period of a month. This was done using a tool that automatically presents Twitter profiles, and allows the recruits to annotate the profile with a classification (bot or human) and add any extra comments. Each account was reviewed by all recruits, before being aggregated into a final judgement using a final collective review (via discussion among recruits if needed).

As well as providing the recruits with the Twitter profile, we also presented summary data to streamline the task. This included: account creation date, average tweet frequency, content posted on user Twitter page, account description, whether the user replies to tweets, likes or favourites received and the follower-friend ratio. We also provide participants with a list of the ‘sources’ used by the account over the month, *e.g.* Twitter app, browser, *etc.* The human workers consider both the number of sources used, and the types of sources used. This

TABLE I  
SUMMARY OF TWITTER DATASET POST-ANNOTATION.

Group	#Bot accounts	#Human accounts	#Bot statuses	#Human statuses
$G_{10M+}$	25	25	70713	77750
$G_{1M}$	295	450	23447	25991
$G_{100k}$	707	740	29777	21087
$G_{1k}$	499	794	16112	5218
<b>Total</b>	1525	2010	140049	130046

is because sources can reveal traces of automation, *e.g.* use of the Twitter API. Additionally, the human worker would also visit a user’s Twitter page and verify the content and URLs posted. Overall, we presented participants with randomised lists that fell into the four popularity groups containing 2500 accounts each. Human annotators were instructed to filter out any account that matched the following criteria: account that does not exhibit activity (*i.e.* no tweet, retweet, reply, and mention), or account that is suspended. In total, the volunteers successfully annotated 3535 accounts. Out of the successfully annotated accounts, 1525 were classified as bots and 2010 as humans. At the time of writing all of the accounts in our dataset were active. Table I provides a summary of the data.

For context, we can cross validate by comparing the agreement of final annotations by the human workers to the BotOrNot annotation. The average inter-annotator agreement compares the pairs of labels by each human annotator to capture the percentage of accounts for which all four annotators unanimously agree. The average agreement is measured as a percentage of agreement, where 0% shows lack of agreement and 100% shows perfect agreement. Our human annotation task shows very high unanimous agreement between human annotators for each popularity group:  $G_{10M+}$  (96.00%),  $G_{1M}$  (86.32%),  $G_{100k}$  (80.66%), and  $G_{1k}$  (93.35%). Whereas, BotOrNot shows lower than average agreement with the final labels assigned by the human annotators:  $G_{10M+}$  (46.00%),  $G_{1M}$  (58.58%),  $G_{100k}$  (42.98%), and  $G_{1k}$  (44.00%). Since, BotOrNot yields a lower accuracy, we restricted ourselves to the dataset of accounts that were manually annotated.

## IV. WHICH MANNERS MAKETH THE BOT?

The purpose of this study is to discover the key account characteristics that are typical (or atypical) of bots and humans. To explore this, we use our data (§III) to empirically characterise bots (dashed lines in figures) and humans (solid lines in figures). To begin, we simply compute the correlation between each feature for bots and humans; Figure 1 presents the results as a heatmap (where perfect correlation is 1.0). It can be immediately seen that most features exhibit very poor correlations (0.0 to 0.35), indicating significant discrepancies between bot and human behaviour — we therefore spend the remainder of this paper exploring these differences in depth.

### A. Account Age

First, we ask *how recently have bots started to participate on Twitter?* To answer this we use account age, which captures the length of time an account has been in operation. Figure

<sup>8</sup>Human annotation task – <http://bit.ly/2ch0YvA>

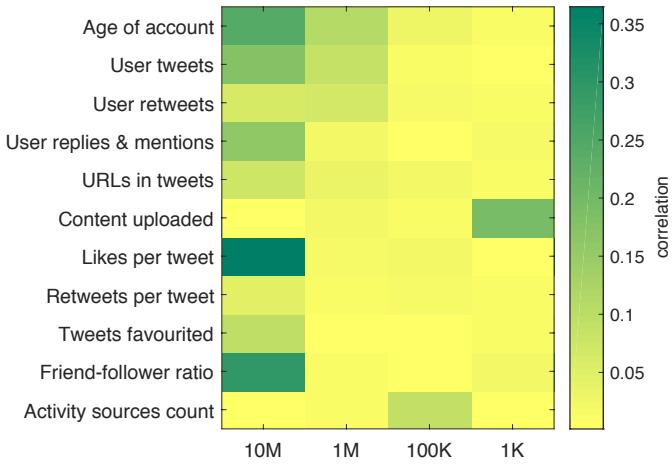


Fig. 1. Spearman’s rank correlation coefficient ( $\rho$ ) between bots and humans per measured metric. The figure shows none (0.0) to weak correlation (0.35) across all metrics, indicating clear distinction between the two entities.

2(a) presents the age of each account, separated into their respective groups (bots *vs.* humans of varying popularities). Unsurprisingly, it can be seen that there is a clear correlation between the age of an account and the number of followers. More noteworthy is the fact that counterpart human accounts are actually slightly ‘younger’ in age than the bot accounts, especially for  $G_{10M+}$ . The oldest bot account is 3437 days old *vs.* 3429 days for the oldest human account – bots are clearly not a new phenomenon. Labels provided by human annotators suggest that these long-standing popular bot accounts have been operated by large reputable organisations as a means of reaching out to their fanbase, viewers and customers, *e.g.* CNN, BBCWorld, NatGeo. In particular, they are interested in spreading news, content, products and services, and are, thus, a very effective tool for these purposes. This further confirms that the Twitter API incentivises the creation and existence of bots on its platform, as highlighted in §I. In fact, the age of some bot accounts suggest that they have long supported the evolution and uptake of Twitter.

### B. Content Generation

Next, we ask *if bots generate more content on Twitter than humans?* Intuitively, one might imagine bots to be capable of generating more content, however, creativity is a clear bottleneck. We initially consider two forms of content creation: a *tweet*, which is an original status written by the account, and a *retweet*, which is repetition of an existing status. When using the term *status*, we are referring to the sum of both tweets and retweets. First, we inspect the amount of content shared by computing the number of statuses (*i.e.* tweets + retweets) generated by each account across the 30 days. As anticipated, humans post statuses less frequently than bots (monthly average of 192 for humans *vs.* 303 for bots), in all popularity groups except  $G_{10M+}$ , where surprisingly humans post slightly more than bots. The sheer bulk of statuses generated by  $G_{10M+}$  (on average 2852 for bots, 3161 for humans in a month) is likely to acquire popularity and new

followers. Overall, bots constitute 51.85% of all statuses in our dataset, even though they are only 43.14% of the accounts.

An obvious follow-up is *what do accounts tweet?* This is particularly pertinent as bots are often reputed to lack original content. To explore this, we inspect the number of *tweets vs. retweets* performed by each account. Figures 2(b) and 2(c) present the empirical distributions of tweets and retweets, respectively, over the 30 days. We see that the retweet distribution is rather different to tweets. Bots in  $G_{1M}$ ,  $G_{100k}$  and  $G_{1k}$  are far more aggressive in their retweeting; on average, bots generate  $2.20\times$  more retweets than humans. The only exception to this trend is  $G_{10M+}$  where humans retweet  $1.54\times$  more often than bots. This is likely driven by the large number of tweets generated by celebrity users — clearly, when tweeting over a hundred times a day, it becomes necessary to rely more heavily on retweeting. That said, typically, humans do generate *new* tweets more often, while bots rely more heavily on retweeting existing content. Generally, humans post 18 tweets for every retweet, whereas bots post 13 tweets for every retweet in all popularity groups except  $G_{10M+}$  (where both entities show similar trends).

Whereas tweets and retweets do not require one-to-one interaction, a further type of messaging on Twitter, via *replies*, does require one-to-one interaction. These are tweets that are created in response to a prior tweet (using the @ notation). Figure 3(a) presents the distribution of the number of replies issued by each account. We anticipate that bots post more replies and mentions given their automated capacity to do so. For  $G_{10M+}$ , bots post only marginally more, on average, than celebrities. This appears to be due to two reasons: bots in  $G_{10M+}$  deploy *chatbots* for addressing simple user queries, and need-based human intervention to engage with followers. Bots in the remaining popularity groups respond twice as frequently as their human counterparts. Again, this seems to be driven by the ease by which bots can automatically generate replies: only the most dedicated human users can compete.

Finally, we briefly inspect the actual content of the tweets being generated by the accounts. We do this using two metrics: number of URLs posted by accounts, and the size of media (*e.g.* pictures) uploaded. Figure 3(b) presents the scatter plot of the number of URLs (*y*-axis) and content uploaded in KB (*x*-axis). Bots place far more external URLs in their tweets than humans (see Table II): 162% in  $G_{10M+}$ , 206% more in  $G_{1M}$ , 333% more in  $G_{100k}$ , and 485% more in  $G_{1k}$ . Bots are a clear driving force for generating traffic to third party sites, and upload far more content on Twitter than humans. Figure 3(c) presents the distribution of the amount of content uploaded by accounts (*e.g.* images). Account popularity has a major impact on this metric. Bots in  $G_{10M+}$  have a  $102\times$  lead over bots in other popularity groups. That said, humans in  $G_{10M+}$  have a  $366\times$  lead over humans in other popularity groups. We conjecture that celebrities have a lot more to share from their lives, but also use uploading as a means to stay connected with their fanbase. Overall, bots upload substantially more bytes than humans too (see Table II): 141% in  $G_{10M+}$ , 975% more in  $G_{1M}$ , 376% more in  $G_{100k}$ , and 328% more in  $G_{1k}$ . Since

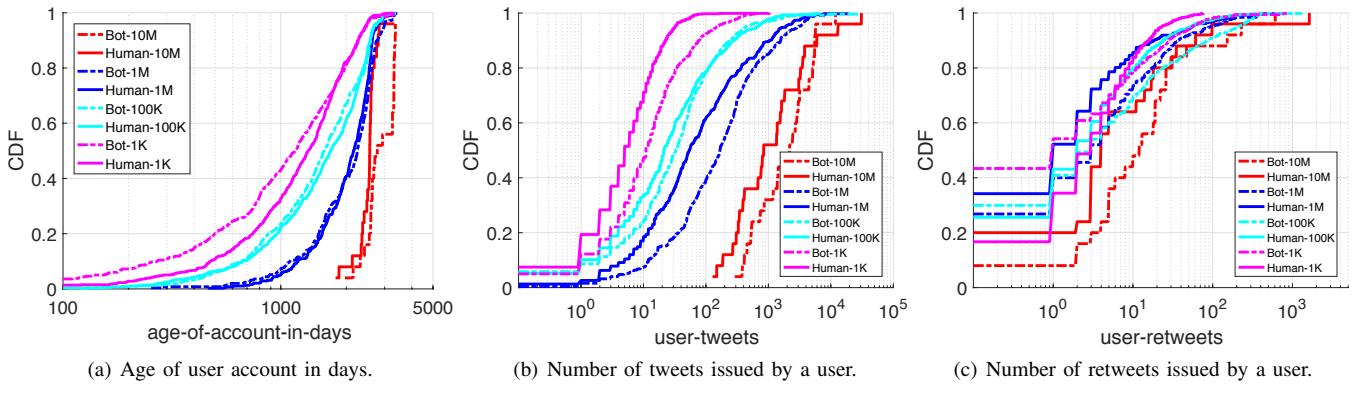


Fig. 2. Account Age; Content Creation: Tweets issued, Retweets issued.

bots in  $G_{10M+}$  mostly belong to news media – sharing news headlines is clearly a means of operating their business.

### C. Content Popularity

The previous section has explored the amount of content generated by accounts, however, this does not preclude such content from being of a low quality. To investigate this, we compute standard popularity metrics for each user group.

First, we inspect the *number of favourites* or *likes* received for tweets generated by the accounts. This is a reasonable proxy for tweet quality. Figure 4(a) presents the empirical distribution of the number of favourites or likes received for all the tweets generated by the profiles in each group. A significant discrepancy can be observed. Humans receive *far* more favourites per tweet than bots across all popularity groups except  $G_{1K}$ . Close inspection by human annotators revealed that bots in  $G_{1K}$  are typically part of larger *social botnets* that try to promote each other systematically for purposes as outlined in §I, whereas human accounts are limited to their social peers and do not usually indulge in the ‘influence’ race. For  $G_{10M+}$ ,  $G_{1M}$  and  $G_{100K}$  popularity groups, humans receive an average of 27×, 3× and 2× more favourites per tweet than bots, respectively.  $G_{1K}$  bots are an exception that receive 1.5× more favourites per tweet than humans. This is a clear sign that the term *popularity* may not be ideally defined by the number of followers. According to the human annotators, this is due to a number of reasons including: humans posting personal news, engaging with their followers, posting less often than bots, staying relevant, posting for their fans (personal things from their life), and not aiming to redirect traffic to external websites. Bots on the other hand, post much more often and do not engage with followers, likely reducing the chances of their tweets being liked.

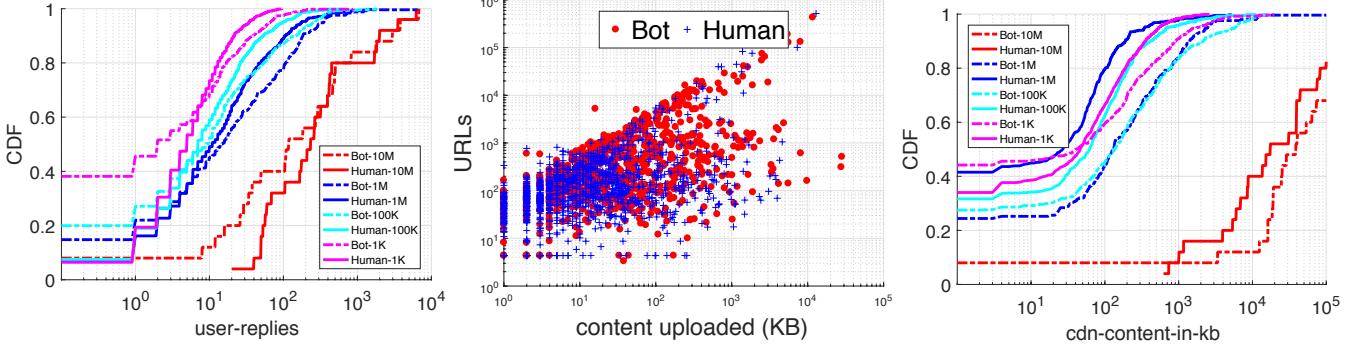
A further sign of content quality is another user retweeting content. Humans consistently receive more retweets for all popularity groups  $G_{10M+}$ : 24-to-1,  $G_{1M}$  and  $G_{100K}$ : 2-to-1, except  $G_{1K}$ : 1-to-1. This difference, shown in Figure 4(b), is indicative of the fanbase loyalty, which is vastly higher for individual celebrities than reputable organisations. In other words, the *quality* of human content appears to be much higher. We can then inspect *who* performs the retweets, *i.e.* do bots tend to retweet other bots or humans? We find that bots

retweeting other bots is over 3× greater than bots retweeting humans. Similarly, humans retweeting other humans is over 2× greater than humans retweeting bots. Overall, bots are retweeted 1.5× more than humans. This indicates a form of homophily and assortativity. To visualise this, Figure 4(c) presents a graph of interactions for the 10M accounts; nodes are the accounts, whilst edges represent a direct interaction *i.e.* retweeted statuses, quoted statuses, replies, and mentions. The above observations of homophily can immediately be seen with bots and humans forming distinct groups. To further confirm this, Figure 5(a) shows the status of interactions if bots are removed from the bot-human interactions (Figure 4(c)). Although interactions are still substantial, they are notably less when bots are excluded. This indicates that, where popularity does exist, it tends to be driven by assortative account types, *e.g.* popular bot accounts are given that status via other bots.

### D. Content Consumption

Whereas the previous metrics have been based on content produced *by* the accounts under-study, our dataset also includes the consumption preferences of the accounts themselves. Hence, we ask *how often do bots ‘favourite’ content from other users and how do they compare to humans?* Intuitively, bots would be able to perform far more likes than humans (who are physically constrained). Figure 5(b) shows empirical distribution of the number of likes performed by each account. It can be seen that, actually, for most popularity groups ( $G_{1M}$ ,  $G_{100K}$ ,  $G_{1K}$ ), humans favourite tweets more often than bots (on average 8251 for humans vs. 5445 for bots across the entire account lifetimes). Linking into the previous discussion, it therefore seems that bots rely more heavily on retweeting to interact with content. In some cases, the difference is significant; for example,  $G_{1M}$  and  $G_{100K}$  see, on average, 2× more likes by humans compared to bots.  $G_{10M+}$ , however, breaks this trend with an average of 1816 by humans compared to 2921 by bots.

We conjecture that there are several reasons for this trend. First, humans have a tendency to appreciate what they like and therefore marking a like is a manifestation of that tendency. In contrast, many bots are simply trying to promote their own tweets and therefore their interactions are based less on likes. We also posit that humans have an incentive to like other



(a) Number of replies and mentions posted by a user.  
(b) Content uploading activity – URLs vs. Content uploaded.  
(c) Size of CDN content in KByte uploaded by a user.

Fig. 3. Content Creation: Replies and mentions, URLs in tweets, Content uploaded on Twitter.

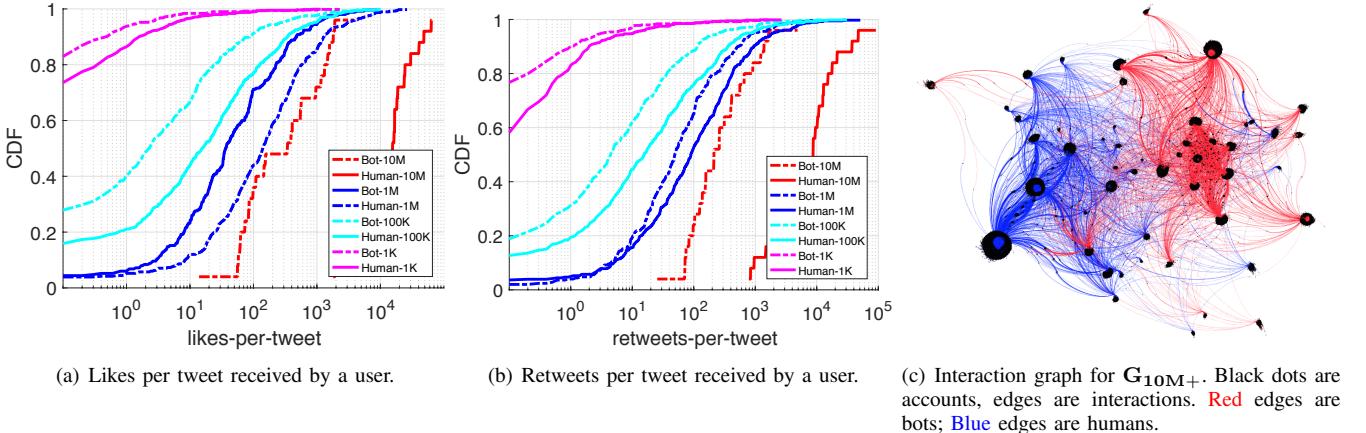


Fig. 4. Content Popularity: Likes per tweet, Retweets per tweet,  $G_{10M+}$  Bots vs. Humans interaction.

tweets, potentially as a social practice (with friends) or in the hope of receiving likes in return [13]. Bots, on the other hand, are less discerning, and feel fewer social obligations. To explore these strategies further, Figure 5(c) plots the number of favourites performed by an account *vs.* the age of the account. It can be seen that more recent (*i.e.* more modern) bots are significantly more aggressive in liking other tweets. Older bots, instead, use this feature less frequently; manual inspection suggests this is driven by the trustworthy nature of older bots, which are largely run by major organisations. In contrast, younger bots adhere less to good social practice and clearly attempt to use favouriting to garner a lot of attention.

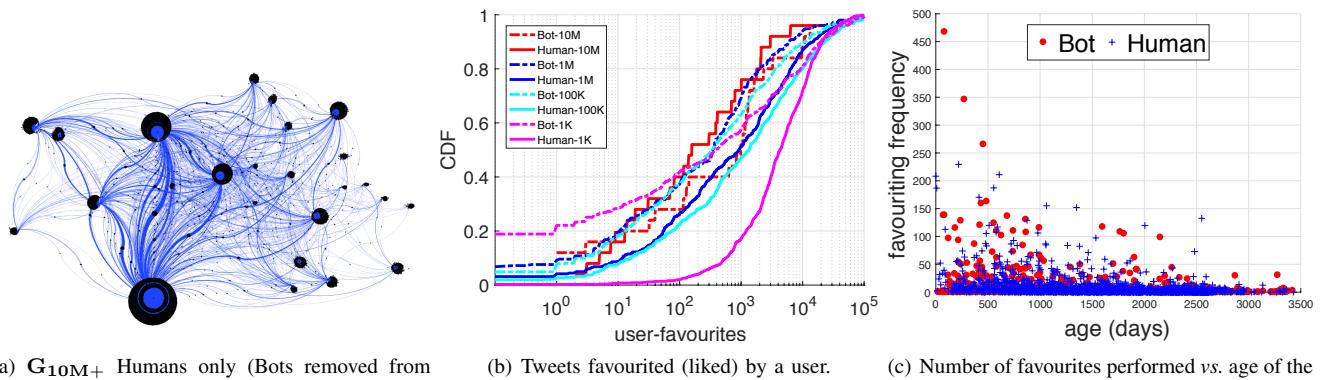
#### E. Account Reciprocity

As well as content popularity, we can also measure reciprocity (*i.e.* friendship). In Twitter, this is classified as a reciprocal follower-relationship (*i.e.* when two accounts follow each other). This is in contrast to a non-reciprocal relationship, whereby an account has many followers who are not followed in return (this is often the case for celebrities). We measure this via the *Follower-Friend Ratio*. Figure 6(a) shows empirical distribution of the *Follower-Friend Ratio* for each group of accounts. The results show that humans display higher levels of friendship ( $G_{10M+}$ : 4.4 $\times$ ,  $G_{1M}$  and  $G_{100k}$ : 1.33 $\times$ ,  $G_{1k}$ : 15 $\times$ ) and thus a lower *Follower-Friend Ratio* than bots.

Previous research [5] argues that humans typically have a ratio close to 1, however, our analysis contradicts this assumption. For celebrities, very popular and mid-level recognition accounts this ratio is in the order of thousands-to-1, irrespective of whether an account is a bot or a human ( $G_{10M+}$ : 629011-to-1 for bots *vs.* 144612-to-1 for humans,  $G_{1M}$ : 33062-to-1 for bots *vs.* 24623-to-1 for humans,  $G_{100k}$ : 2906-to-1 for bots *vs.* 2328-to-1 for humans). In fact, even the ratios for low popularity accounts are not 1, but consistently greater ( $G_{1k}$ : 30-to-1 for bots *vs.* 2-to-1 for humans). This is caused by a human propensity to follow celebrity accounts (who do not follow in return), as well as the propensity of bots to indiscriminately follow large numbers of other accounts (largely in the hope of being followed in return).

#### F. Tweet Generation Sources

Finally, we inspect the tools used by bots and humans to interact with the Twitter service. This is possible because each tweet is tagged with the *source* that generated it; this might be the website, app or tools that employ the Twitter API. Figure 6(b) presents the number of sources used by human and bot accounts of varying popularities. Surprisingly, it can be seen that bots actually inject tweets using more sources than humans (*cf.* Table II). This is unexpected as one might expect bots to use a single source (*i.e.* an API or their own

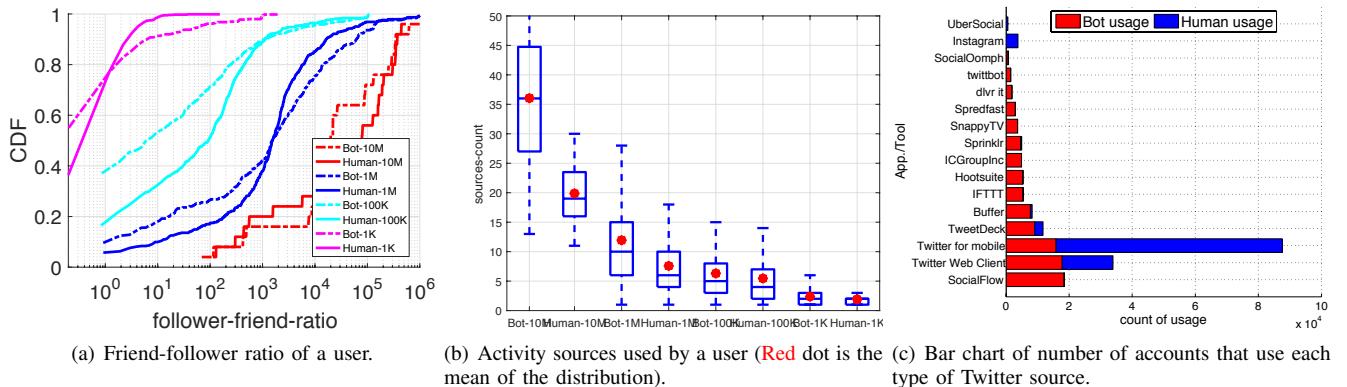


(a)  $\mathbf{G}_{10M+}$  Humans only (Bots removed from Fig. 4(c)).

(b) Tweets favoured (liked) by a user.

(c) Number of favourites performed vs. age of the account.

Fig. 5. Content Popularity:  $\mathbf{G}_{10M+}$  Humans only; Content Consumption: Likes performed, Favouriting behaviour.



(a) Friend-follower ratio of a user.

(b) Activity sources used by a user (Red dot is the mean of the distribution).

(c) Bar chart of number of accounts that use each type of Twitter source.

Fig. 6. Account Reciprocity; Tweet Generation Sources: Count of Activity Sources, Type of Activity Sources.

automated tool). However, paid third party services exist that can be utilised for the automation purpose.

To explore this further, Figure 6(c) presents the number of accounts that use each source observed. It can be seen that bots use a multitude of third-party tools. Bot news services (especially from  $\mathbf{G}_{10M+}$ ) are found to be the heaviest users of social media automation management and scheduling services (*SocialFlow*, *Hootsuite*, *Sprinklr*, *Spredfast*), as well as a Cloud-based service that helps live video editing and sharing (*SnappyTV*). Some simpler bots (from  $\mathbf{G}_{100k}$  and  $\mathbf{G}_{1k}$  groups) use basic automation services (*Dlvr.it*, *Twittbot*), as well as services that post tweets by detecting activity on other platforms (*IFTTT*). A social media dashboard management tool seems to be popular across most groups except  $\mathbf{G}_{1k}$  (*TweetDeck*). Interestingly, it can also be seen that bot accounts regularly tweet using the Web/mobile clients — it seems likely that these accounts have a mix of automated and human operation. In contrast, 91.77% of humans rely exclusively on the Web/mobile clients. That said, a small number (3.67%) also use a popular social media dashboard management tool (*TweetDeck*), and automation and scheduling services (*Buffer*, *Sprinklr*). This is particularly the case for celebrities, who likely use the tools to maintain high activity and follower interaction — this helps explain the capacity of celebrities to so regularly reply to fans (§IV-B). These things should be noted when using the tweet sources to identify bots.

## V. DISCUSSION, CONCLUSIONS & FUTURE WORK

Bots exercise a profound impact on Twitter, as this paper thoroughly investigates. Table II summarises key differences between bots and humans. For each popularity group the table shows which feature most suitably distinguishes which entity, *i.e.* bot ( $\mathcal{B}$ ) or human ( $\mathcal{H}$ ). The magnitude of differences is represented by \* (considerable) and \*\* (large). Table II also includes statistical significance (*signif.*) of the mean *difference* between humans and bots across all popularity groups, as measured by the *t-test*: at 99% confidence level (extremely different), and 95% (very different).

This confirms a number of noteworthy behavioural differences between bots and humans in this characterisation study. For example, confirming past work, we observe a significant and longstanding bot presence on Twitter (§IV-A). We show that bot users in  $\mathbf{G}_{1M}$  tweet more often than humans in the same category. However, there are a small number of features for which humans and bots are more similar, particularly in  $\mathbf{G}_{10M+}$ , where we find humans often exhibit bot-like activity (*e.g.* tweeting a lot).

That said, overall, we find that bot accounts generate larger amounts of tweets than their human counterparts, whilst they (proportionally) rely far more heavily on retweeting existing content and redirecting users to external websites via URLs (§IV-B). These trends appear to be driven by their difficulty in creating novel material when compared to humans. Despite

TABLE II

FEATURE INCLINATION:  $\mathcal{B}$  IS MORE INDICATIVE OF BOTS, WHEREAS  $\mathcal{H}$  IS MORE INDICATIVE OF HUMAN BEHAVIOUR, AND  $\circ$  IS NEUTRAL (*i.e.* BOTH EXHIBIT SIMILAR BEHAVIOUR). \* REPRESENTS MAGNITUDE OF INCLINATION: \* IS CONSIDERABLE DIFFERENCE, \*\* IS LARGE DIFFERENCE. *signif.* SHOWS STATISTICAL SIGNIFICANCE OF EACH FEATURE AS MEASURED BY *t-test*.

Feature & value	Fig.	10M+	1M	100K	1K	signif.
Older age of account in days	2(a)	$\mathcal{B}$	$\circ$	$\circ$	$\circ$	
More user tweets	2(b)	$\circ$	$\mathcal{B}^*$	$\mathcal{B}^*$	$\mathcal{B}^*$	
Higher user retweets	2(c)	$\mathcal{H}^*$	$\mathcal{B}^*$	$\mathcal{B}^*$	$\mathcal{B}^*$	99%
More user replies and mentions	3(a)	$\circ$	$\mathcal{B}^*$	$\mathcal{B}^*$	$\mathcal{B}$	99%
More URLs in tweets	3(b)	$\mathcal{B}^{**}$	$\mathcal{B}^{**}$	$\mathcal{B}^{**}$	$\mathcal{B}^{**}$	99%
More CDN content (KByte) uploaded	3(c)	$\mathcal{B}^{**}$	$\mathcal{B}^{**}$	$\mathcal{B}^{**}$	$\mathcal{B}^{**}$	95%
Higher likes received per tweet	4(a)	$\mathcal{H}^{**}$	$\mathcal{H}^{**}$	$\mathcal{H}^{**}$	$\mathcal{B}$	99%
Higher retweets received per tweet	4(b)	$\mathcal{H}^{**}$	$\mathcal{H}^{**}$	$\mathcal{H}^{**}$	$\mathcal{B}$	99%
More tweets favoured (liked)	5(b)	$\mathcal{B}^{**}$	$\mathcal{H}^{**}$	$\mathcal{H}^{**}$	$\mathcal{H}^{**}$	99%
Higher follower-friend ratio	6(a)	$\mathcal{B}^{**}$	$\mathcal{B}^*$	$\mathcal{B}^*$	$\mathcal{B}^{**}$	
More activity sources	6(b)	$\mathcal{B}^*$	$\mathcal{B}$	$\mathcal{B}$	$\mathcal{B}$	99%

the voluminous amount of tweets contributed by bots, we find that they still fail to beat humans in terms of *quality* though, *e.g.* humans receive a mean of  $19\times$  and a median of  $41\times$  more likes per tweet than bots across all popularity groups (§IV-C). The divergence between retweet rates is even greater. Humans receive a mean of  $10\times$  and a median of  $33\times$  more retweets per tweet than bots.

Regardless of their reputed sophistication, it seems that bots still fail to impress users when viewed through the lens of these metrics. Bots also seem to spend less time liking others' tweets (§IV-D); in other words, humans see greater value in showing their appreciation of other content and users. This trend is mirrored in other metrics too, *e.g.* humans tend to have a higher proportion of reciprocal relationships (friendships) than bots (§IV-E). When combined, we believe that these propensities are heavily influenced by a form of social obligation that bots clearly do not feel. The lack of such considerations may lead to dramatic changes in social structures and interactions in the longterm (as the bot population increases).

It was also particularly interesting to note the similarities between celebrity users and bots; we even found that some celebrities use bot-like tools (§IV-F). Finally, we found the two user groups exhibit homophily both in terms of one-hop interactions and wider communities. For example, bots retweeting other bots is over  $3\times$  more regular than bots retweeting humans, whereas humans retweeting other humans is over  $2\times$  greater (§IV-C).

There is a series of interesting and open questions that need to be explored and answered in our future work. A major line of future work is expanding the size of our data to understand how bots impact information dissemination more widely. Currently, we only focus on the accounts that actively disseminate content (*i.e.* tweets or retweets). It would be worthwhile to also inspect less active bots and humans, *e.g.* fewer than 1K followers. Moreover, we wish to go beyond bot *vs.* human studies to understand the *purpose* of the (bot)

accounts, *e.g.* news disseminators, socio-political infiltrators or activists, marketing bots, spam and malicious content propagators, *etc.* It would be interesting to further explore the content creation process and perform topical analysis, *i.e.* how and what material do bots tweet and which approaches gain most traction, especially for bots that create original content. Finally, our investigations have revealed inaccuracies in existing bot detection algorithms (hence our use of manual classification). Thus, utilising our findings to fine-tune an automated but reliable bot detection tool is a major remaining challenge. These topics are reserved as our future research, and plan to extend our measurement platform accordingly to help us seek answers.

## REFERENCES

- [1] N. Abokhodair, D. Yoo, and D. W. McDonald. Dissecting a social botnet: Growth, content and influence in twitter. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 839–851. ACM, 2015.
- [2] L. M. Aiello, M. Deplano, R. Schifanella, and G. Ruffo. People are strange when you're a stranger: Impact and influence of bots on social networks. In *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2012.
- [3] Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu. The socialbot network: When bots socialize for fame and money. In *Proceedings of the 27th Annual Computer Security Applications Conference, ACSAC '11*, pages 93–102, New York, NY, USA, 2011. ACM.
- [4] D. Boyd, S. Golder, and G. Lotan. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*, pages 1–10, Jan 2010.
- [5] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia. Who is tweeting on twitter: Human, bot, or cyborg? In *ACSAC '10*. ACM, 2010.
- [6] S. Cresci, R. D. Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi. The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race. *CoRR*, abs/1701.03017, 2017.
- [7] C. A. Davis, O. Varol, E. Ferrara, A. Flammini, and F. Menczer. Botornot: A system to evaluate social bots. In *Proceedings of the 25th International Conference Companion on World Wide Web, WWW '16 Companion*, pages 273–274, 2016.
- [8] N. Dugué, A. Perez, M. Danisch, F. Bridoux, A. Daviau, T. Kolubako, S. Munier, and H. Durban. A reliable and evolutive web application to detect social capitalists. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, pages 741–744. ACM, 2015.
- [9] C. Edwards, A. Edwards, P. R. Spence, and A. K. Shelton. Is that a bot running the social media feed? testing the differences in perceptions of communication quality for a human agent and a bot agent on twitter. *Computers in Human Behavior*, 33:372 – 376, 2014.
- [10] C. Freitas, F. Benevenuto, S. Ghosh, and A. Veloso. Reverse engineering socialbot infiltration strategies in twitter. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, pages 25–32. ACM, 2015.
- [11] Z. Gilani, L. Wang, J. Crowcroft, M. Almeida, and R. Farahbakhsh. Stweeler: A framework for twitter bot analysis. In *Proceedings of the 25th International Conference Companion on World Wide Web, WWW '16 Companion*, pages 37–38, 2016.
- [12] K. Lee, B. D. Eoff, and J. Caverlee. Seven months with the devils: A long-term study of content polluters on twitter. In *ICWSM*, 2011.
- [13] L. Scissors, M. Burke, and S. Wengrovitz. What's in a like?: Attitudes and behaviors around receiving likes on facebook. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, CSCW '16*, pages 1501–1510, New York, NY, USA, 2016. ACM.
- [14] C. Wagner, S. Mitter, C. Körner, and M. Strohmaier. When social bots attack: Modeling susceptibility of users in online social networks. *Making Sense of Microposts (# MSM2012)*, 2, 2012.
- [15] X. Wu, Z. Feng, W. Fan, J. Gao, and Y. Yu. *Detecting Marionette Microblog Users for Improved Information Credibility*, pages 483–498. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.