A Typology of Socialbots

Gregory Maus
School of Informatics and Computing
Bloomington, Indiana
gmaus@iu.edu

Onur Varol
School of Informatics and Computing
Bloomington, Indiana
ovarol@indiana.edu

ABSTRACT

Socialbots continue to rise in prominence as a new type of organism in humanity's social ecosystem, put to use in everything from government manipulation of public opinion to novel personal entertainment. As important a force as socialbots are, their detection and collective analysis has been complicated by their diversity. Based on an in-depth analysis of academic and journalistic literature examining a huge array of known social bots, we present a multi-tiered set of traits intended to classify all types of bots. The taxonomy establishes a lexicon for discussing and comparing bots, as well as refining detection methods based upon distinctive characteristics of each bot type. One trait that has proven especially important for detection is the way in which bots network with each other, as overtly connected bot networks can be discovered en masse together when anomalies from a few individuals are noted. Conversely, bots in the same campaign without overt connections can be detected together due to improbable similarities in behavior. This schema also classifies bots via the dimensions of alleged humanity, degree of autonomy in several vectors, operator, and objectives. By clarifying discussion of socialbot traits, this paper intends to facilitate communication among researchers investigating the full diverse span of socialbot phenomena

KEYWORDS

socialbots, social bots, online automation, digital propaganda, twitterbot, social botnet, social spam, spam 2.0

ACM Reference format:
Gregory Maus and Onur Varol. 2017. A Typology of Socialbots. In *Proceedings of ACM Web Science conference, Troy, NUY, June 2017 (WebSci17)*, 8 pages.
DOI: 10.475/123_4

1 INTRODUCTION

Socialbots represent a new organism in humanityfis social ecosystem. To put that into perspective, consider that most recent time anything comparable happened was 10,000 years ago, when we allowed cats into our homes. Like cats, bots have proven manipulative, stealthy, unpredictable, difficult to control, and nearly ubiquitous on the internet. Unlike (most) cats, however, bots have also been instrumental in oppressing civic dissent, spreading disinformation, facilitating mass fraud, and fomenting civil unrest. The purpose of this analysis is thus to provide a classification framework for facilitating discussion of these creatures, distinguishing harmless domestic pets from dangerous predators, and advancing our ability to identify the lions stalking an unsuspecting public.

For purposes of this paper, fisocialbotsfi refers to automated or largely automated programs that interface with online platforms in largely the same way that a typical human would be expected to: they hold normal accounts, make connections, and post content. This definition thus excludes automated callers, smartphone-based chat/assistant bots, and other programs meant to interact with users entirely offline or via specialized interfaces. However, it includes a broad range of both malign and benign systems across social media platforms. Itfis impossible to assess with certainty how many socialbots are currently active. Not only do many bots work to conceal their nature, but most social media companies prefer not to let independent researchers investigate such matters for several reasons, including privacy concerns for their users and fears that findings of high bot usage may invalidate their claims of subscriber numbers and thus harm financial valuations.

However, forthcoming algorithmic evaluations estimate that 9-15% of Twitter accounts are bots ([TBA], 2017) in contrast to Twitterfis own estimates of less than 5% [1]. Even after Instagram purged millions of suspected bot accounts researchers estimated that 8% of the sites accounts are bots.[1] Socialbots have been found attempting to covertly influence politics as early as the 2010 U.S. midterm elections [2]. Since then they have been used by governments and political parties around the world to demoralize and disrupt dissent, spread government messages (overtly and covertly), and inflate the follower counts of political figures.[3] Government opposition figures have also retaliated with their own socialbot armies. [4][5] Most recently, a study concluded that approximately a fifth of Twitter users discussing the 2016 US presidential election might have been bots.[6]

Bots have been regularly linked to misinformation campaigns.[6][7][2] Given how even small changes in initial social media ratings can dramatically impact the long-term reach of content[8] their impact should not be underestimated.

Bots have been found fraudulently boosting reviews for mobile apps[9], ad-clicks[10], video views [11]. Theyfive been tied to phishing campaigns on Facebook[12], Twitter[13], Craigslist [14], 2016), and Tindr [15].

Bots can be manipulative in other ways. Infidelity-oriented dating site Ashley Madison operated tens of thousands of bots masquerading as real female users in order to entice its male users to buy the sitefis paid services. Internal company reports concluded that 80% of men who purchased the services were motivated by contact with one of these bots.[16]

On the bright side theyfire supportive in administrative tasks on

WebSci17, Troy, NUY
© 2017 ACM. 123-4567-24-567/08/06…$15.00
DOI: 10.475/123_4

Wikipedia [17], are useful for curating information [18] [19], and can provide a number of other unique services such as algorithmically created music on demand [20], reporting the arrival of dictators in Geneva[21], and even provided custom-generated sympathetic messages[22].

2 RELATED WORK

Previous work in bot classification schema has been relatively sparse.

Oentaryo et al. (2016) propose a broad behavioral categorization of bots into broadcasters (which share benign content with general audiences), consumers (which aggregate content from multiple sources and/or provide updates), and spammers (which post malicious or irrelevant content).[18]

Mitter et al. (2013) offer a multi-level categorization of malicious social bot campaigns based on the campaignfis targets, vulnerabilities exploited, bot account type, attack method, and result .[23] While thorough and detailed in analyzing malicious bot attacks, it spares little analysis for the bots themselves and none for benign bots.

Woolley (2016) divides the use of political bot campaigns by governments into the three categories of demobilizing opposition, fostering pro-government messages, and padding follower numbers for government and/or candidate accounts.[3]

Ji et al (2016) provides an overview of mechanisms used by bots to evade detection.[24]

3 CATEGORIZATION CAVEATS

Attempting to categorize socialbots runs into similar issues as categorizing biological organisms in that new discoveries may complicate or invalidate existing classification systems. New discoveries of socialbots are likely to occur in two ways.

Firstly, new types of bots can be developed with novel applications and features that defy or expand known categories.

Secondly, existing covert socialbots are discovered with some regularity.[25][26][27][6][28][29][5][12][30][31][11][16][32][4][33][34][3]. Thus we know that many bots trying to evade detection exist, but we can't know how many exist that successfully avoid detection. Therefore there could already be unknown legions of socialbots, some of which might defy current categorization.

More broadly, the categories are in some cases fuzzy, especially in terms of Objectives as bots may have multiple objectives and the distinctions between them can be somewhat blurry.

4 CLASSIFICATION DIMENSIONS

4.1 Alleged Humanity

This dimension encompasses the extent to which a bot attempts to pass itself off as a human and the nature of the profile information that it might use to support this claim.

4.1.1 Honest Bot. These bots make no attempt to hide their nature. Their descriptions or other profile features openly out them as such. Their behavior may not be even intended to remotely resemble that of a human.

4.1.2 Intentionally Ambiguous. Customer support bots and others designed to appear human to only a casual user, but without making any claims as such and makes little effort to hid its nature.

4.1.3 Fleshmask. Fleshmasks deceptively and consistently portray themselves as humans. It is difficult to imagine circumstances in which this is positive for all involved.

Generic: Generics have few features that stand out, often lacking a profile picture or description text.

Imposter Imposters are intended to deceive other users into believing that the bot is a specific real human being. An enigmatic instance that might have involved a network of such imposters was reported in 2009: fake Facebook profiles for fimore than 100 scientists, policy-makers and journalists many linked to stem-cell research, whose identities have been purloined to create a convincing fi!? but bogus fi!? network of apparent friendsfi the ultimate purpose and instigator of which seems unresolved .[52] Though the degree to which the network was automated remains unclear, a paper presented on the same month as the above report demonstrated a means for automated sequential identity theft and impersonation on social media.[41]

A definitely automated example is that of the Twitter bot which used the name and photo of investigative journalist Jon Ronson, much to the journalist's annoyance. His interview with the bot's creators suggests that they intended it as a commentary on the nature of identity online and automation.[42]

Contacting the individual in question through known, secure channels and asking them about the account would be the most effective way to verify whether it is a bot imposter. As of early 2016 Facebook was reportedly testing tools to automatically detect potential imposters and ask users for verification.[53]

Technically, this category also includes so-called "cyborgs", bots that supplement a user's work under their own names with the user's knowledge and consent, like posting pre-approved content on a regular schedule[44]. After all, the bots are taking action presumed to be the user's own.

Near-Imposter Near-Imposters copy account information almost wholesale, perhaps with slight variations on the name or other features to prevent duplication. This is presumably intended to create consistency among the elements of the account. A 2015 study identified over 15,000 pairs of such near-imposters on Twitter, though it would be impossible to verify that that they are all bots.[30] Comparable bots have been identified on Instagram.[31] The extreme similarity between the account and that of another user can be the telltale identifier. The aforementioned study developed an machine learning-based algorithmic identifier built upon analyzing the similarities between accounts.[?]

Chimera Chimeras pull traits from many different sources: a profile picture from google images, a description from another user, etc.

Reviewers might identify poorly stitched together chimeras from inconsistencies between profile features, such as a description stating that theyfire a life-long resident of Arizona while birthplace is listed as Calcutta. Searches for identical information on content from other users or stock databases might also uncover telltale overlaps with other accounts.

Table 1: Typology Graph

Dimension	Type	Sub-Type		Examples
Alleged Humanity	Honest Bot			[21][35][36][37][38]
	Intentionally Ambiguous			[39]
	Fleshmask	Generic		
		Chimera		[16][40]
		Near-Imposter		[30][31](Jimbots)
		Imposter		[41][42]
		Custom		[43]
Operator	Individual			[39][44][38]
	Government			[3][25][26][45][34][5]
	NGO			[21][36][41][27][6]
	Commercial	Loyalist		[16][35]
		Mercenary	Promiscuous	[27](Jimbots)
			Single-Client	[16]
Bot Coordination	Hive			[25][26][46][41][40]
	Aspen			[16][32][43]
	Singleton			[21][39][35][36][44][38]
Operator Transparency	Covert			[39][25][27][26][45][32][46][41][16][43](Jimbots)
	Overt			[21][35][37][44][38]
Objectives	Novel Content Broadcasting			[21][35][44][38]
	Content Promotion			[25][27][26][32][33][46](Jimbots)
	Content Curation			
	Information Acquisition			[41][13][12]
	Channel Disruption			[25][34][45]
	Network Graph Alteration			[41][36](Jimbots)
	Transaction			[47][48][49][50]
	Entertainment			[37][39][38]
	Proof of Concept			[35][46][36][39][43]
	Metric Manipulation			[10][16][51][9][11]
	Motivator			[16][43]

Custom Custom fleshmasks are manually designed, presumably by a human being, though an advanced means for automatically generating sophisticated, unique accounts could be quite valuable and so can not be ruled out as a future or existing application.

4.2 Operator

The type of group supporting the bot. This may also be, but isnfit necessarily the creator and designer of the bot.

4.2.1 Individual. These bots are run by individuals for a wide variety of reasons including hobbyism and personal amusement.

4.2.2 Government. These bots are operated by states. In some cases the distinction between government-operated and political party-operated bots can be blurry or non-existent.

4.2.3 NGO. As in other cases, NGO is a broad category encompassing political parties, charitable organizations, academic groups, and terrorist organizations.

4.2.4 Commercial. Commercial operators are any number of organizations seeking monetary profit. Considering the areas of legal grey currently associated with bots, this includes aboveboard corporations, criminal operations, and everything in-between.

Loyalist Loyalists are directly intended to contribute to the profitability of their operators.

Mercenaries Mercenaries are operated by for-profit groups or individuals on behalf of clients.

Promiscuous Mercenaries Promiscuous mercenaries promote the interests of multiple distinct clients, which can result in an incongruous variety of distinct causes supported, perhaps even contradictory ones or those in an improbable variety of languages.

Single Client Mercenaries Single client mercenaries are operated by for-profit groups on behalf of only a single client, perhaps in order to avoid the telling lack of identity coherence typical of promiscuous mercenaries, which can lead to them being implausibly single-issue focused.

4.3 Bot Coordination

Socialbots may coordinate with other soci bots in order to achieve outcomes that are only possible through multiple accounts.

4.3.1 Hive. Hives network with eachother directly via Friending, Following, Liking, or similar connections to reinforce each otherfis credibility. Hives vary in the degree of how tightly packed they are, that is the degree to which they follow/retweet/etc. each other. Some prior literature ha referred to such networks as fibotnetsfi,[46][40]

but this framework specifically avoids that term in order to avoid confusion with the earlier digital security meaning of the term.
Tightly clustered hives tend to be identifiable through recurrent clusters of direct echoing, as in the improbably strong retweeting relationship within the Syrian Social Botnet[25].
In constrast, looser hives can be identified through improbably similar echoes despite the lack of overt connections between the accounts. For a blatant example, during the 2016 U.S. election reporters identified many apparently Hispanic Twitter users who tweeted (not retweeted) identical pro-Trump messages within seconds of each other without an obvious relationship between them.[27]
The two approaches can also be used together, as in the case of documenting the network of nearly 20,000 pro-Kremlin bots. The researcher identified the use of identical phrases (and telltale identical errors) among seemingly disparate accounts, followed their in-group relations to map out the network, then confirmed their bot-hood by noting improbable account similarities and creation times in the vast network.[26]
Hives tend to be identified together, such that in finding one the reviewer need only look at its connections in order to find the others. Less sophisticated or less adventurous networks may be more insular due to an inability or lack of attempt to get others to reciprocate their networking requests.

4.3.2 Aspen. Like the tree, aspens may seem to be distinct entities on the surface, lacking even the indirect overt connections of loose hives, but are secretly united in purpose. Presumably this is intended to reduce the chances of discovery for all bots in the network.
Even though they are not directly connected with each other, they may produce or promote suspiciously similar content. The account information may also be formulaic across the network. More subtly, networks of spambots in YouTube comments have been algorithmically identified through linguistic analysis to find telltale variations on the same essential construction.[32]

4.3.3 Singleton. As might be guessed, singletons are not part of a larger network of bots either overtly or covertly.

4.4 Operator Transparency

This refers to whether the bot is publicly tied to its operator or not.

4.4.1 Overt. The bot is publicly claimed as a representative of the organization operating it, or being operated on its behalf. This can be the case even if its nature as a bot is unclear, as in many customer service bots.

4.4.2 Covert. The relationship between the operator and the bot is distinctly opaque.

4.5 Content Creation Autonomy (Spectrum)

This spectrum reflects the degree to which an automated program produces the content to be broadcast. On one extreme are bots that perfectly mirror the content provided to them by a human being. Slightly more autonomous would be bots those that add minor random variation on content (perhaps so as to avoid detection.)[32] Still more autonomous creates entirely novel content based upon the request of the operator or other users.

4.6 Content Broadcast Autonomy

Subtly distinct from creation autonomy, this is about the selection which content to broadcast (regardless of origin), as well as when and where to share it. With the least autonomy, they are instructed manually by the operator as to what and how they broadcast. Slightly more autonomy would be to allow variations on time or audience, perhaps to obscure their relationship, or to select content from an established list.
Higher autonomy could allow for the bot to choose the audience, time, and source of the content it broadcasts, perhaps including content not originating with itself or the operator.

4.7 Network Graph Autonomy (Spectrum)

This is about how much freedom the bot has to choose how it engages with the network whether that be other users directly or established content. No autonomy suggests that the bot couldnfit choose its own Friends/Followers, etc. or even react to such requests. Limited autonomy would allow the bot to react to connection requests automatically. More would allow it to reach out to other users or automatically engage with content.

4.8 Objectives

The objectives of operatorsfi bots can be fuzzy and are certainly not restricted to a single category. It seems probable that bot developers will continue finding new applications for bots. As artificial intelligence continues to advance, botsfi creators may give them broader license in deciding how to pursue more generalized objectives than is currently feasible. One could imagine, for example, a sophisticated bot intended to use advanced sociological analytics in order to autonomously decide how to promote the ideology of the operatorfis choice rather than just promoting selected content, but that would seem to beyond current capabilities. Thus, this list cannot hope to be exhaustive.

4.8.1 Novel Content Broadcasting. These bots are based around broadcasting unique content. Versions with high content creation autonomy algorithmically generate new content sometimes based on user input (as in a Twitterbot that composes music based on Tweeted specifications [37], sometime automatically (such as the bot that Tweets algorithmically generated fantasy maps every hour [38].)
Lower autonomy bots may simply broadcast content from a preestablished list, as in the ”cyborgs” that post on a schedule from a selection of content created by a user (sometimes with repetition)[44]. This may shade into Content Promotion (below) if it includes linking to content from elsewhere.

4.8.2 Content Promotion. Promoters echo or draw attention to content found elsewhere. The most direct example would be Sharing/Retweeting/reposting/etc. content from elsewhere. Just slightly less direct would be up-voting/Liking/etc. content or Following/Friending/etc. channels/accounts to increase visibility. The content need not be on the same site, as with url link spam. The methods can also be highly indirect, exemplified by bots that apparently exploited how Google indexed Tumblr for black hat search engine optimization by liking many benign Tumblr pages in order to tie them to backlinks on the botsfi offsite pages [33].

4.8.3 Content Curation. Essentially what Oentaryo et al. refer to as ficonsumption botsfi [18], curators are something of an inverse of promoters, searching for relevant content and then post it and/or links to it unaltered in an established format and/or platform at the operatorfis request. While its output may be accessible to others aside from the operator, it is distinguished from promoters in that it does not actively seek to promote the content outside the curatorfis own channel of consenting users.

The line between curators and novel content producers could conceivably become somewhat blurry if the curator adds analysis or modifications to the content it queries.

4.8.4 Information Acquisition. Socialbots can be used to collect a wide variety of information including general platform aggregates, public account information, semi-public account information (that information only available to Friends and such), hidden account information, or non-account information (presumably obtained through conversation with the individual.)// A 2011 8-week study infiltrating Facebook with a network of socialbots for collecting usersfi Friends-only profile information achieved a 59.1% acceptance rate of Friend requests. [28]// Phishing is also included in this category.

4.8.5 Channel Disruption. Channel disruptors are intended to compromise a communication channel. Pro Kremlin operatives seem to have pioneered the use of bots to flood dissentersfi Twitter hashtags and render them unusable during the disputed 2011 Russian election[45] [3], a technique mimicked to drown out the Mexican protest hashtag #YaMeCanse

[34][29]. A variation also seems to have been used by Pro-Assad Syrian socialbots, diluting the Syria tweets with news unrelated to the Syrian civil war and in some cases entirely unrelated to Syria.[25]

4.8.6 Network Graph Alteration. Bots could also be used to influence the overall graph of the social network working to alter usersfi connections to each other or tying content together in ways that previously werenfit present. The simplest example of this would be the many types of bots designed to autonomously Follow/Friend/etc. users in order to connect with them. More advanced versions might in-turn try to connect the users with other targeted users (or bots)fi!?or perhaps work to undermine existing connections.

An example is the Botivist, a Twitterbot that searches Twitter for those who have recently expressed a strong opinion on a given social issue and then connects (via calls to action and discussion) them with others who have similarly expressed opinions on it, achieving a 80% reply rate with the most effective strategy tested.[36]

4.8.7 Transaction. These bots conduct a variety of financial transactions.

Perhaps the most infamous of these are ticket or auction "snipers." In what's been described as an "epidemic" bots buy hundreds or thousands of event tickets within seconds of their online availability, which the bots' masters then resell at a mark-up for resale at markups.[47] In 2013 bots were estimated to represent 90% of the U.S. traffic on online vendor Ticketmaster.[48] Despite being illegal in some states, these bots continue to drive up ticket prices and have earned the personal enmity of textitHamilton creator Lin-Manuel Miranda.[54] Sniping bots have also been seen in use for last-second auction bidding, restaurant reservations, and plane tickets.[48]

Bots that masquerade as human users in online gambling and bots used for acquiring resources in MMO's would be considered more sophisticated forms of transaction bots. Gambling bots have been shown capable of beating world-class human players in poker tournaments [55] and have made in-roads in gambling sites, despite crack-downs[49]. In some cases, the bots seem to be run by the sites themselves.[50]

Bots have also been persistently endemic in acquiring virtual goods in online games, often for resale.[56][57]

Stretching the definition of socialbots, high speed trading algorithms could also be considered an example.

The recent flurry of development in chatbots for commerce and customer service is thus far mostly focused upon specialized apps and interfaces [58], and thus are technically not socialbots by this paperfis definition, but it seem little stretch to imagine that at some point that bots integrated similarly to normal human social platform users could be used as intermediaries for transactions or customer service. Even if not, the close relatedness of such programs merits a mention.

4.8.8 Entertainment. Broadly, the antics of bots may be intended as entertainment. This could be for those it interacts with, as with many novelty content creators. Alternatively, it could be intended entirely to entertain the operator and perhaps those told of it, as with many trolling bots.

4.8.9 Proof of Concept. Some bots are intended either for experiments or to demonstrate a conclusion.

These experiments can sometimes have unexpected and unpleasant results, as when Microsoftfis learning chatbot Tay was corrupted by its Twitter conversations into making racist and misogynistic Tweets after just 24 hours of operation [35].

4.8.10 Motivation. Motivators work to encourage or discourage specific behaviors, generally behaviors within the platform on which they exist.

Bots used for manipulating users into purchasing premium dating accounts and bots intend to decrease racist harassment are both covered in the examples below.

Admittedly, this definition can become blurry. For example, political bots promoting content in favor of their candidates could be considered motivators for voting a certain way.

4.8.11 Metric Manipulation. Metric manipulators influence quantitative metrics, usually to inflate perceived popularity with humans or certain demographics.

Bots for fraudulently boosting ad click-throughs[10] would be one of the purer expressions of this type, as the purpose is to give the expression of human interest to the ad-space purchaser, without attempting to influence humans to actually use the content.

For demographic-specific manipulation, bots are frequently used on dating sites to balance out the apparent gender ratio according to industry insiders[51], though as in the case of Ashley Madison, they can also pursue additional objectives if they interact with users[16].

App review boosting [9](and other review boosting) would also be an example, as it is intended to nudge content recommendation sorting algorithms to give them higher priority, but this also shades into content promotion, as the end goal is to bring the attention of humans. Video view fraud[11] is a similarly mixed case.

Finally, bots intended to boost followers, shares, etc. might be considered to have tangential elements of this as well (along with content promotion and possibly network graph alteration), if the purpose is to simply increase the aggregate of such figures.

5 ILLUSTRATIVE EXAMPLES

These examples are intended to showcase a diverse sampling of some of the more well-documented socialbots and how they are classified by this schema.

The disproportionate representation of Twitter-based bots is an unfortunate artifact of that platform's transparency (and thus ease of identifying/observing bots) relative to other platforms.

5.1 Jimbots: Covert Near-Imposter Promiscuous Mercenary Hive of Content Promoters/Network Graph Alterers

Named for a Twitter account holder whom the hive almost impersonated twice (originalhandle @JimKeplinger copied to @iJmKeplingre and @JimKeplirgen, along with his old description and picture, but slight changes to his place of residence) the Jimbots follow a select few others from their own network along with sites from their apparent clients. Tweets from the bots suggest bilingualism and improbably diverse interests including lesbian pornography, RT Espanol, Drudge Report, Puerto Rican baseball, and (fittingly) quantitative marketing analytics.

Further analysis of the Jimbot network will be presented in a future publication.

5.2 Ashley Madison's "Angels": Covert Chimera Loyalist/Single-Client Mercenary Aspen of Metric Manipulators/Motivators

Infidelity-centered dating site Ashley Madison admitted to having used widespread automated fake accounts on their site posing as women.[59] Not only did the bots' presence bolster the site's alleged male:female ratio[51], but it seems they were key to enticing users to purchase the premium services as internal company reports concluded that 80% of men who purchased premium accounts were motivated by contact with one of these bots[16].Internally nicknamed "Angels", these tens of thousands of accounts were created by hand by both employees and contractors (hence why some would be considered loyalists and the others mercenaries)[60], largely using content recycled from the site's abandoned accounts[16]. The bots seem to have had numerous flaws, such as the fembots engaging with gay men,[61] or occasionally message other bots[16]. One user who filed a legal complaint also noted that bots from different accounts would send him identical messages and would all keep unusually regular schedules of login times regardless of holidays.[16] (Note how, as is typical of aspens, the bots were identifiable by improbable similarities in behaviors.) According to anonymous reports, customers who complained about these unfaithful practices and demanded refunds were blackmailed into silence.[62] Other dating sites have also made use of bots, as demonstrated by the Federal Trade Commission's 2014 settlement against a conglomerate that owned 18 dating sites and used bots to deceive users into paying for upgraded accounts.[63] As of 2015, industry insiders alleged that fake accounts continued to be a common practice for dating sites.[51]

5.3 GVA Dictator Alert: Overt Honest NGO Singleton Novel Content Broadcaster

Created and maintained by Swiss journalist Franois Pilet and his cousin Julien Pilet, the GVA Dictator Alert is a Twitterbot that determines and announces when a plane registered to an authoritarian government arrives at and departs from Geneva. The intent is to shed light on potential dictator transactions with Swiss financial services. The bot synthesizes information from existing amateur aircraft spotters based on air traffic control signals that communicate the unique tail numbers of planes and cross-references with databases of planes publicly registered to authoritarian governments.[21] According to the project's site, information from the bot was used in 2016 to open a case by Geneva public prosecutors against the Vice President of Equatorial Guinea for money laundering, eventually resulting in prosecutors seizing 11 luxury cars owned by the VP.[64]

5.4 Tweetment Bots: Covert Custom Fleshmask Individual Aspen Proof of Concept/Motivators

The Tweetment Bots were created as a study to analyze the effectiveness of social admonishments for discouraging racist harassment on Twitter. The bots varied in terms of apparent race (based on name and their cartoon portrait) and (purchased) follower count in order to gauge differences in impact based on these factors. When the researcher identified Twitters using a specific anti-black racist term, the researcher would manually check to see if it indeed constituted racist harassment by a white male adult and, if so, mark the user for engagement by one of the bots. The bot would then chide the user for such language. Admonishments by allegedly white, high follower bots had an apparent significant affect decreasing future use of racist terminology by the users, but low-follower allegedly white bots and all allegedly black bots seemed to have little, if any affect.[43]

5.5 Arguetron: Covert Intentionally Ambiguous Singleton Proof of Concept/Entertainer

Designed by activist Sarah Nyberg, this bot tweets generic progressive-leaning talking points approximately every ten minutes. It never targets anyone with the tweets and thus the many alt-right Twitter users who tried to angrily engage with it must have specifically sought it out for argument/harassment by searching for tweets on its topics. The account has many red flags which should have caused it to stand out as a bot: a handle of "@arguetron" (though its given name was "Liz"), an improbably regular tweet schedule, and the fact that all of its replies to comments directed at it were generated randomly from a generic calm response library without any actual relation as to what was said to it. Despite these obvious shortcomings, people would argue with or harass the bot for hours. Nyberg intended the project "to expose the hypocrisy of the sorts of people who say fifeministsfi and fisocial justice warriorsfi are hypersensitive. Even just mild statements of fact can have them absolutely freak out fi!? and end up sending abuse, even fi!? to a bot that responds calmly and just explains over and over again that they're wrong." The fact that the arguments have gone so long despite Arguetron's responses being unrelated to her interlocutors' statements also supports her conclusion that "So many arguments, especially on a place like Twitter, are almost content-neutral."(Hence why it qualifies as a Proof of Concept.) [39] It also seems that she's derived entertainment from watching the arguments and sharing them with others.

5.6 Syria Social Botnet: Covert Fleshmask (type unclear) Government Hive of Channel Disruptors/Content Promoters

The Syrian Social Botnet, identified and documented by a Seattle team in April-December 2012, was a network of approximately 130 Twitterbots posting in both English and Arabic to advance narratives favorable to the Bashar al-Assad regime in Syria and dilute the Syria hashtag with information unrelated to the ongoing Syrian civil war. "Core bots" (115/130 accounts in the network) at the center of the web would tweet roughly every 6 minutes and retweet the content of other core bots (though some core bots never retweeted at all.) "Peripheral bots" (some of which, the researchers concluded, may have actually been legitimate users) would then occasionally retweet the cores. Most of the tweeted content was news sourced from a narrow group of outlets sympathetic to Assad. A strikingly large portion (31.8%) were associated with Syria, but made no mention of the conflict tearing the country apart in start contrast to other accounts tweeting about Syria. The researchers hypothesized that this was intended to obscure tweets related to the war. The researchers identified the hive by finding clusters of users who retweeted each other about Syria with improbable regularity. After finding that 17 accounts had been suspended that followed a specific news aggregator bot, they snowballed the regular retweet connections to map out the network.[25] It should be stated that the Syrian Social Botnet is not the largest pro-government bot hive that's been found, merely the most thoroughly documented. In 2015, researcher Lawrence Alexander uncovered a network of nearly 20,000 pro-Kremlin Twitterbots.[26]

5.7 The Star Wars Botnet: Covert Chimera Hive (operator type and objective unknown

This wretched hive of 350,000 twitterbots is the largest ever documented–over .1% of the entire active Twitter population (though the researchers state that they've found an even larger network of 500,000 bots.) Despite its size it managed to lay dormant for three years without detection. It earns its name from the fact that all of its content is random quotations from Star Wars novels, often cut off mid-word, and sometimes with random hash tags inserted. The researchers were first tipped off by the fact that many tweets were being geo-tagged in areas that were uninhabited, both desert wastes and oceans. Indeed it would turn out that the network was evenly distributed over simple rectangles covering the U.S. and Europe with no relation to actual population centers. Additionally, the bots all selected "Twitter from Windows Phone" as their source, despite this only being used by a tiny minority of the general Twitter population. These factors, combined with the disproportionate use of certain words (such as "jedi") compared to the general population, allowed them to train an algorithm to identify the full network. The researchers theorize that the network went unnoticed for so long by defying accepted wisdom for typical bot behavior: they only tweeted infrequently, roughly mimicked human language by tweeting from novels, never used urls in their tweets (a tactic of spammers), had relatively realistic profiles (in correspondence Echeverria mused that they used amalgamations of user names, but was uncertain), and only had a small group of friends. The purpose and creators of the network are still unknown, but the researchers theorize that it was intended to potentially be reactivated at a later date or sold.[40]

6 CONCLUSION

By creating a standardized typology of socialbots we intend to clarify understanding and discussion of socialbot phenomena in all its diversity.

Already the clarity granted by this delineation of bot categories has inspired specific new approaches in our research group for bot identification methods and refining our BotOrNot detection system. We consider this evidence of the classification schema's current effectiveness and hope that other researchers will similarly find it useful.

That said, portions of the typology will remain works in progress. The current messiness of the Objectives is rather unsatisfying and it remains to be seen if such untidiness is a fundamental necessity given the potential variety of uses to which bots might be put or if it can be streamlined. Certainly it's difficult to imagine how the objective dimension could be dispensed with entirely without losing significant understanding of the bot itself.

Ultimately, regardless of how the classification schema evolves, or even if it is entirely superseded by another one, the conceptual clarity of such systems will be vital for understanding our new digital companions and competitors as they continue to advance and exercise ever greater social influence (whether we notice it or not.)

7 ACKNOWLEDGEMENTS

Thank you to Fil Menczer and John Paolillo for their advice and feedback.

Classification of the Syrian Social Botnet and the Star Wars Botnet would also not have been possible without the helpful correspondences of Norah Abokhodair and Juan Echeverria Guzman respectively.

REFERENCES

[1] Deepa Seetharaman. Fake accounts still plague instagram despite purge, study finds. June 2015.

[2] Jacob Ratkiewicz, Michael Conover, Mark Meiss, Bruno Goncalves, Alessandro Flammini, and Filippo Menczer. Detecting and tracking political abuse in social media, 2011.

[3] Samuel Woolley. Automating power: Social bot interference in global politics. *First Monday*, 21(4), 2016.

[4] Sergey Sanovich, Denis Stukal, Duncan Penfold-Brown, and Joshua Tucker. Turning the virtual tables: Government strategies for addressing online opposition with an application to russia. June 2015.

[5] Michelle Forelle, Phil Howard, Andrés Monroy-Hernández, and Saiph Savage. Political bots and the manipulation of public opinion in venezuela. *CoRR*, abs/1507.07109, 2015.

[6] Alessandro Bessi and Emilio Ferrara. Social bots distort the 2016 u.s. presidential election online discussion. *First Monday*, 21(11), 2016.

[7] Dan Misener. Political bots spread misinformation during u.s. campaign. November 2016.

[8] Maria Glenski, Thomas J. Johnston, and Tim Weninger. Random voting effects in social-digital spaces: A case study of reddit post submissions. *CoRR*, abs/1506.01977, 2015.

[9] Zhen Xie and Sencun Zhu. Appwatcher: Unveiling the underground market of trading mobile app reviews. In *Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, WiSec '15, pages 10:1–10:11, New York, NY, USA, 2015. ACM.

[10] Adrian Neal, Sander Kouwenhoven, and OB SA. Quantifying online advertising fraud: Ad-click bots vs humans. Technical report, 2015.

[11] Miriam Marciel, Rubén Cuevas, Albert Banchs, Roberto Gonzalez, Stefano Traverso, Mohamed Ahmed, and Arturo Azcorra. Understanding the detection of fake view fraud in video content portals. *CoRR*, abs/1507.08874, 2015.

[12] Hongyu Gao, Jun Hu, Christo Wilson, Zhichun Li, Yan Chen, and Ben Y. Zhao. Detecting and characterizing social spam campaigns. In *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, IMC '10, pages 35–47, New York, NY, USA, 2010. ACM.

[13] Sidharth Chhabra, Anupama Aggarwal, Fabricio Benevenuto, and Ponnurangam Kumaraguru. Phi.sh/$ocial: The phishing landscape through short urls. In *Proceedings of the 8th Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference*, CEAS '11, pages 92–101, New York, NY, USA, 2011. ACM.

[14] George Khourey. How to spot bot-generated craigslist scams. November 2016.

[15] Better Business Bureau. Scam alert – how to spot spam profiles on tinder, July 2015.

[16] Annalee Newitz. How ashley madison hid its fembot con from users and investigators. September 2015.

[17] Maxime Clment and Matthieu J. Guitton. Interacting with bots online: Usersfi reactions to actions of automated programs in wikipedia. *Computers in Human Behavior*, 50:66 – 75, 2015.

[18] Richard J. Oentaryo, Arinto Murdopo, Philips K. Prasetyo, and Ee-Peng Lim. *On Profiling Bots in Social Media*, pages 92–109. Springer International Publishing, Cham, 2016.

[19] A marvellous incomplete compendium of reddit automatons bots, author=Lock, Duncan, year=2013, month=June, day=19, url=http://duncanlock.net/blog/2013/06/19/a-marvellous-incomplete-compendium-of-reddit-automatons-bots/.

[20] Ngai Zhang. The musical twitter bot: Who has the copyright for ai-facilitated works?, October 2016.

[21] Amar Toor. This twitter bot is tracking dictators' flights in and out of geneva. October 2016.

[22] Victoria Turk. This bot tries to empathise with your emotional tweets. September 2016.

[23] Silvia Mitter, Claudia Wagner, and Markus Strohmaier. A categorization scheme for socialbot attacks in online social networks. *CoRR*, abs/1402.6288, 2014.

[24] Yuede Ji, Yukun He, Xinyang Jiang, Jian Cao, and Qiang Li. Combating the evasion mechanisms of social bots. *Computers Security*, 58:230 – 249, 2016.

[25] Norah Abokhodair, Daisy Yoo, and David 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.

[26] Lawrence Alexander. Social network analysis reveals full scale of kremlin's twitter bot campaign. April 2015.

[27] Natalie Andrews. Pro-trump twitter bots at center of nevada mystery. February 2016.

[28] Yazan Boshmaf, Ildar Muslukhov, Konstantin Beznosov, and Matei 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.

[29] Klint Finley. Pro-government twitter bots try to hush mexican activists. August 2016.

[30] Oana Goga, Giridhari Venkatadri, and Krishna P. Gummadi. The doppelgänger bot attack: Exploring identity impersonation in online social networks. In *Proceedings of the 2015 ACM Conference on Internet Measurement Conference*, IMC '15, pages 141–153, New York, NY, USA, 2015. ACM.

[31] Adrianne Jeffries. It's your face. it's your photos. meet the creepiest kind of instagram spambot. September 2014.

[32] Derek O'Callaghan, Martin Harrigan, Joe Carthy, and Pádraig Cunningham. Network analysis of recurring youtube spam campaigns. *CoRR*, abs/1201.3783, 2012.

[33] Ernie Smith, Seth Millstein, Chris Tognotti, Patrick deHahn, Scott Craft, and Matthew Keys. Why you might get tons of tumblr likespam (and what to do about it), October 2011.

[34] Pablo Suárez-Serrato, Margaret E. Roberts, Clayton A. Davis, and Filippo Menczer. On the influence of social bots in online protests. preliminary findings of a mexican case study. *CoRR*, abs/1609.08239, 2016.

[35] James Vincent. Twitter taught microsoftfis ai chatbot to be a racist asshole in less than a day. March 2016.

[36] Saiph Savage, Andres Monroy-Hernandez, and Tobias Höllerer. Botivist: Calling volunteers to action using online bots. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, pages 813–822, New York, NY, USA, 2016. ACM.

[37] LnH Music. Lnh music, 2016.

[38] Margaret Rhodes. This bot tweets a totally fantastical map every hour. August), day = 16, url = https://www.wired.com/2016/08/bot-tweets-totally-fantastical-map-every-hour/ 2016.

[39] Kaitlyn Tiffany. The internetfis alt-right are mistakenly arguing with a bot. October 2016.

[40] Juan Echeverría and Shi Zhou. Thestar wars' botnet with¿ 350k twitter bots. *arXiv preprint arXiv:1701.02405*, 2017.

[41] Leyla Bilge, Thorsten Strufe, Davide Balzarotti, and Engin Kirda. All your contacts are belong to us: Automated identity theft attacks on social networks. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, pages 551–560, New York, NY, USA, 2009. ACM.

[42] Jon Ronson, Lucy Greenwell, and Remy Lamont. Jon ronson v 'jon ronson' spambot ... part two - video.

[43] Kevin Munger. Tweetment effects on the tweeted: Experimentally reducing racist harassment. *Political Behavior*, pages 1–21, 2016.

[44] Craig Timberg. As a conservative twitter user sleeps, his account is hard at work. February 2017.

[45] Brian Krebs. Twitter bots drown out anti-kremlin tweets. December 2011.

[46] J. Zhang, R. Zhang, Y. Zhang, and G. Yan. On the impact of social botnets for spam distribution and digital-influence manipulation. In *2013 IEEE Conference on Communications and Network Security (CNS)*, pages 46–54, Oct 2013.

[47] Eric T Schneiderman. Obstructed view: What's blocking new yorker's from getting tickets, 2016),url=https://ag.ny.gov/pdfs/Ticket$_{S}ales_{R}eport.pdf.$

[48] Cal Flyn. The bot wars: why you can never buy concert tickets online. August 2013.

[49] Gabriel Dance. Poker bots invade online gambling. March 2011.

[50] Robert Blincoe. Masters of the poker face. February 2009.

[51] Caitlin Dewey. Ashley madison faked female profiles to lure men in, hacked data suggest. 2015.

[52] Brian Krebs. Fake facebook pages spin web of deceit. April 2009.

[53] Karissa Bell. Facebook is testing a feature that alerts you if someone is impersonating your account. March 2016.

[54] Lin-Manuel Miranda. Stop the bots from killing broadway. June 2016.

[55] Nick Abou Risk and Duane Szafron. Using counterfactual regret minimization to create competitive multiplayer poker agents. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: Volume 1 - Volume 1*, AAMAS '10, pages 159–166, Richland, SC, 2010. International Foundation for Autonomous Agents and Multiagent Systems.

[56] Nathan Grayson. World of warcraft bot maker calls it quits after massive ban wave. May 2015.

[57] Seong Hoon Jeong, Ah Reum Kang, and Huy Kang Kim. Analysis of game bot's behavioral characteristics in social interaction networks of mmorpg. In *Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication*, SIGCOMM '15, pages 99–100, New York, NY, USA, 2015. ACM.

[58] Michael Quoc. The state of bots: 11 examples of conversational commerce in 2016. June 2016.

[59] David Z. Morris. Ashley madison used chatbots to lure cheaters, then threatened to expose them when they complained. 2016.

[60] Annalee Newitz. The fembots of ashley madison. 2015.

[61] Annalee Newitz. Ashley madison code shows more women, and more bots. 2015.

[62] Jose Pagliery. Ashley madison threatened to expose customers who disputed bills. 2016.

[63] FTC. Online dating service agrees to stop deceptive use of fake profiles. October 2014.

[64] GVA Dictator Alert. About, 2016.