This is a pre-copy-editing, author-produced PDF of an article accepted following peer review for publication in Computers, Environment and Urban Systems

The Role of Bots in U.S. Real Estate Development Online Communication

Authors
Justin Hollander*, Ruth Potts, Maxwell Hartt, Minyu Situ, and Alex Seto
*Author for correspondence

Introduction

Social media has provided a digital landscape in which citizens can share their opinions on decision-making in real-time, ask questions, and actively participate in discussions around urban issues. The wide adoption of social media over the past two decades has resulted in a surge of public interest in using social media to engage with others around key place-based issues. In short, social media has changed the way in which people engage with planning. Internet-based participatory tools provide a valuable opportunity for policymakers and planners to engage a wide variety of stakeholders and inform decision-making processes at a low cost. However, low cost and high potential engagement comes with a risk. The anonymity and digitization of participatory processes also created an opportunity for subversion from groups with alternate and possibly nefarious interests. Social media has created a new mechanism through which citizens can express opposition to or support for planning projects. This raises concerns regarding whether social media accounts could be used to unfairly influence participatory planning processes, and amplify certain voices within such processes.

Internet-enabled communities have been highlighted as an opportunity to foster better discussions and debates around community visioning and planning (Mitchell 1995; Cowley and Hollander 2010; Hollander, et al. 2016). Framed as a logical extension of the seminal works of communicative planning scholars like Forester (1999) and Innes and Booher (2010), urban research has focused on ways that the political processes of shaping places can benefit from online community engagement and dialogue. The prominent concerns regarding the movement of community planning processes onto online platforms, which has largely occurred as anticipated, have focused on the digital divide and the lack of human touch with online exchanges. Discussion of the risks associated with using online platforms to engage with the public has largely ignored the potential manipulation of social media in local planning processes for the benefit of certain groups.

While there is a growing body of literature exploring the methods and impacts of social media manipulation on political discourse, no research has considered the potential damage that such manipulation can and does cause in the local planning context. This project examines the vulnerability of local planning processes to social media manipulation.

We aim to answer the following research questions in order to guide our observation of the roles and risk of artificial intelligence power in online communities:

1) What are the threats of automated social media accounts on public discourse?
2) How common are these automated social media accounts in real estate development?
3) What is the scale and scope of the threat of these automated social media accounts to the real estate development process?

The paper begins with an exploration of the academic literature around social media and potential manipulation strategies and risks. The literature findings are then used to ground our empirical research measuring and analyzing the risks associated with social media manipulation in local planning processes.

Social media and planning practices

Social media enables users to create and rapidly share information through internet-enabled mobile devices and computers, which considerably shortens the time between events occurring and the news being shared on social media. Social media enables users to create and respond to information in real-time, enabling users to adapt their communications based on constantly emerging information (Tsou et al., 2013). Since their emergence in the mid-2000s, social media has been increasingly used by urban planners to engage with different publics and stakeholders on urban issues. Like other digital participation methods such as public participation geographic information systems (PPGIS), theorists and practitioners have mostly emphasised the potential benefits of social media platforms for community engagement, and enabling planning authorities to give feedback to the community on planning issues in real time. Where PPGIS is seen as a means of enabling citizen 'scientists' to collectively map spatial information and contribute actively to decision-making, social media is framed as more of a conduit for individuals to share information and perspectives on specific issues (Chmielewski et al., 2018; Hasanzadeh et al., 2020; Williamson & Ruming, 2020). The planning literature particularly emphasises the usefulness of social media platforms such as Facebook, Twitter, and online forums (Evans-Cowley & Griffin, 2011; Fredericks & Foth, 2013; Williamson & Parolin, 2013; Wilson et al., 2017). Social media is purported by scholars to enable planners greater opportunities to build social capital (Afzalan & Evans-Cowley, 2015; Mandarano et al., 2010), create stronger relationships with their local communities (Arribas-Bel et al., 2015; Firmstone & Coleman, 2015; Hanzl, 2007), and better share information (Evans-Cowley & Conroy, 2006; Fredericks & Foth, 2013; Williamson & Parolin, 2012). Social media is also described by scholars as a natural evolution of participatory and communicative planning concepts espoused by Healey (2003) and Innes (1995).

Social media is a relatively novel method of engaging the public for urban planners, compared to many of the more traditional face-to-face methods employed in local planning processes. Social media as a public engagement tool gives the public a greater amount of power than traditional planning processes to decide which issues are discussed, how issues are discussed, the planning issues they engage with irrespective of their spatial location, and the extent to which their identity is shared (Evans-Cowley, 2010; Wilson et al., 2017). Social media also offers planners a number of opportunities to engage with a wider variety of citizens across a broader spatial area, at a low cost (Evans-Cowley, 2010; Wilson et al., 2017). As a result, since the mid-2000s many local planning authorities and departments have expanded their participation through newly established social media accounts, particularly on Twitter and Facebook (Firmstone & Coleman, 2015; Fredericks & Foth, 2013; Williamson & Parolin, 2012). Simultaneously, communities have embraced social media as a means of connecting with like-minded citizens and expressing their concerns on local planning issues (Evans-Cowley, 2010). This means that social media is increasingly and concurrently being used by both citizens and government planning authorities/departments to engage with planning issues (Evans-Cowley, 2010; Williamson & Ruming, 2015). These two approaches are discussed below.

Citizens, Social Media and Planning Issues

Social media has become highly popular within the community as a means of engaging with local planning authorities on urban issues (Ertiö & Bhagwatwar, 2017; Lin & Geertman, 2019). Platforms such as Facebook have largely been embraced because they enable citizens to self-organise into 'groups' with common interests, rather than the more open structure of more anonymous platforms such as Twitter (Evans-Cowley & Hollander, 2010). Studies of such groups and their engagement with planning issues have found that they often form to resist development proposals or plans (Evans-Cowley & Hollander, 2010; Williamson & Ruming, 2015). Local planning authorities are also often unaware of or deliberately disengaged with citizen-run Facebook groups opposing development, due to their lack of structure and formal linkages to local planning processes (Evans-Cowley & Hollander, 2010).

Citizen and community use of social media to engage with local spatial issues and planning processes is uneven. At its most basic level, citizen engagement may involve ‘following’ or ‘liking’ a group, page, or key spokesperson with an emphasis on a particular local planning issue or spatial area. On the other hand, some citizens may engage at a higher level, and post personal comments on such groups/pages, share content with other citizens, or organise in-person activities. For example, in one study citizens in a Finnish community self-organised a ‘Pop-up Cleaning Day’ via a Facebook group in response to concerns about waste and the sustainability of consumption patterns (Horelli et al., 2015). Similarly, citizens in Sydney (Australia) created a Facebook group as part of their opposition to a proposed master plan for a local precinct (Williamson & Ruming, 2015).

Government Planning Authorities and Social Media

Similar to citizens, local planning authorities/departments’ use of social media to engage with local planning issues is uneven (Evans-Cowley & Conroy, 2006; Williamson & Parolin, 2012). Despite the potential for social media enabling heightened levels of dialogue with the community and greater discussion between citizens of local planning issues in real time, actual use of social media by planning authorities have often been limited to use as a ‘top down information dissemination channel’ (Fredericks & Foth, 2013, p. 246). This approach to using social media such as Twitter and Facebook involves planning authorities/departments providing citizens with information related to the timing/location of face-to-face consultation events, and upcoming projects in the local area, as opposed to more genuine explorations or dialogue between citizens around specific local planning issues (Evans-Cowley & Conroy, 2006; Williamson & Parolin, 2012).

Scholars have explained the limited applications of social media for public engagement by planners as reflecting concerns of political implications, privacy, and confidentiality (Fredericks &

Foth, 2013), high time cost of moderating and managing social media accounts (Afzalan & Evans-Cowley, 2015), as well as a lack of skills or knowledge regarding how to effectively use the data and types of information generated from social media in planning processes (Evans-Cowley & Griffin, 2011; Williamson & Parolin, 2013). In addition to this, some studies have consistently argued that planners are often wary of new technologies and thus considered 'slow adopters'(Klosterman, 1997; Klosterman & Landis, 1988), suggesting that social media use by planners may increase in time as they become more comfortable with its parameters.

Social media and the Risks

The risks of using social media in planning processes are not well studied. Case studies in the planning literature highlight challenges associated with the accuracy of communication between citizens around planning issues on social media and concerns about the distortion of the facts (Afzalan & Evans-Cowley, 2015; Evans-Cowley & Griffin, 2011; Williamson & Parolin, 2013). Afzalan and Evans-Cowley (2015, p. 80) found social media generated 'an environment in which distortions were propagated' through poor communication, despite the presence and active participation of a planner within a forum focussed on green infrastructure in Oregon, USA. In this instance, the distortion of ideas was not the result of intentional deception as warned by Williamson and Parolin (2013), however does demonstrate the risk of miscommunication and misrepresentation of ideas. Evans-Cowley (2010) argue that the degree to which citizens can present themselves anonymously may limit the degree to which planners know who exactly they are interacting with online, and thus may create an environment where deception is more likely.

The risks associated with social media use are well explored in the digital communications literature. One of the primary arguments within this literature is that the nature of social media platforms (cheap, accessible, live, and limited moderation) makes them vulnerable to manipulation (Paquet-Clouston et al., 2017; Wald et al., 2013; Woolley & Howard, 2018). As a result, users of such platforms may easily misrepresent themselves and key ideas, and intentionally influence public discourse around major political or social issues. There is substantial evidence to suggest that this has already occurred and swayed key political debates globally, including the 2016 Brexit debate (Bastos & Mercea, 2018), 2016 US presidential debate (Bessi & Ferrara, 2016), and the 2015 conflict between Ukraine and Russia (Hegelich & Janetzko, 2016). While manipulation of social media can reflect individual users with nefarious intentions, it is most frequently associated with 'social bots'.

Online Community: Twitter

One of the most popular platforms of social media, Twitter was launched in 2006 as an online micro-blogging service (Tsou et al., 2013). Twitter users can write status messages to their 'followers' on Twitter, of which those original written messages are called 'tweets' and those re-shared ones are called 'retweets' (Twitter, 2010). Abundant tweets have provider researchers an opportunity to better to understand online human communication (Miller 2011) and behaviours (Perreault and Ruths 2011)**.**

Ch'Ng (2015) analyzed Twitter hashtags to explore out the forming pattern of Twitter communities. Ch'Ng (2015) found that users of Twitter frequently self-organised around a common ideology or common judgement of ideology. This suggests that users can form communities relationally, reinforcing perspectives and creating cohesion between members through online discussions and defending against the conflict with other users or communities (Ch'Ng, 2015). According to Ch'Ng (2015) individuals are key influencers on Twitter, with ideas proliferating and being amplified by groups and increasing the extent to which other Twitter users are exposed and influenced themselves. The use of social media interactions such as Retweeting the posts of other users, may enable citizens to feel like their voice is contributing to a broader conversation.

Twitter is perhaps more vulnerable to manipulation than Facebook due to its more open format, whereby any individual can engage in broader debates irrespective of the connections between them. This risk to planning processes is associated with the potential for information to be manipulated by other users or intentional interventions. This in turn could shift sentiment among entire communities on Twitter, and potentially mislead public decisions that rely on social media for public engagement. Thus, how to identify original ideas and their source becomes a key issue when using specific social media platforms to engage with citizen perspectives.

Social Bots and Planning Practice

The U.S. Senate Intelligence Committee accused the Russian Government in May 2018 of engaging in cyberattacks intended to disrupt the 2016 presidential election. Journalists and policy-makers documented many of these attempts to sway political discourse towards Donald Trump particularly through Twitter and Facebook, and raised broader concerns that online communities are at risk of being manipulated by actors with nefarious intentions (U.S. Senate Select Committee on Intelligence, 2017). While these issues are well discussed at the national political scale, there has been no discussion of how manipulation may influence politics or planning decisions at the local scale.

‘Social bots’ have become more prevalent and sophisticated in their interactions on social media in recent years (Alarifi et al., 2016). ‘Social bots’ are computer programs that can mimic human behaviour and perform basic activities on social media, such as sharing or posting content, and sending friend requests. Case studies have shown that ‘social bots’ are the primary source of social media discourse manipulation, rather than individual users (Alarifi et al., 2016; Wald et al., 2013). Social bots enable third parties to ‘reach and potentially influence a large and diverse population of web users’(Boshmaf et al., 2011, p. 93) beyond the capabilities of an individual user. While some bots are transparent regarding their non-human nature (e.g. Emoji Weather USA - a weather sharing bot on Twitter), others are deliberately misleading and seek to represent themselves as human users (Edwards et al., 2014). There are a number of strategies used by 'social bots' to manipulate users and discourse summarised in Table 1. These strategies are rarely used independently and can be combined to more effectively spread rumours, false information, and reinforce specific perspectives (Wald et al., 2013). 'Social bots' are increasingly difficult to identify and distinguish from human users due to their increasing sophistication and computer learning, making it difficult for human users of social media to discern trustworthy sources of information from unreliable sources (Alarifi et al., 2016).

Table 1: Strategies used by social bots.

Strategy	Description	Sources
Astroturfing	Creation of multiple accounts for sharing or resharing opinions to suggest greater consensus on an issue than actually exists.	(Badawy et al., 2018; Ratkiewicz et al., 2011)
Twitter bombs / Smoke screening	Flooding a hashtag used by opposing arguments/positions with counter content, deliberately obscuring one party's arguments with the opposition's counter arguments.	(Abokhodair et al., 2015; Stieglitz et al., 2017)
Tagging influential users	Mentioning 'high profile' or influential users in posts containing low quality content to exploit the credibility of the follower and gain the attention of their large follower base.	(Shao et al., 2018; Subrahmanian et al., 2016)
Spamming	Publishing a large number of similar posts in a short period of time.	(Ferrara, 2018; Stieglitz et al., 2017)

A number of international studies have examined the role of 'social bots' in influencing elections internationally (Howard et al., 2017; Neudert et al., 2017). They used open-source tools such as BotOrNot (Badawy et al., 2018) to identify bots, and compared the portion of reliable to unreliable URLs, comparing the twitter traffic on certain days, and looking at the geospatial spread of users that have interacted with bots. The studies found that it was possible to identify bots and distinguish reliable and unreliable sources of information on Twitter. ‘Social bots’ tended to exaggerate news and label themselves as ‘satirical’ while reliable sources had clear evidence linked within their articles.

Data and Methods

Online communities intersect closely with the geographic places where people live, work, and play. The forums, hashtags, and other community markers that appear to be aspatial online actually connect with real places during community planning processes. In this study we measured the presence of bots in numerous large-scale real estate development projects in the U.S., across a wide range of geographies, uses, size, and controversy levels.

Employing a convenience sampling strategy, we first culled major urban planning and real estate development professional journals, magazines, and websites (including the American Planning Association’s *Planning* magazine and the Urban Land Institute’s *Urban Land* magazine) for the names and locations of large-scale real estate development projects being proposed or developed over the last three years. Our goal was to find approximately 20 such projects across the following dimensions: 1) type of development (e.g. residential, mixed-use, sporting facility, etc.), 2) size (e.g. square feet and/or acreage), and 3) region of the USA. Only projects that have generated three or more social media groups, blogs, or websites by non-developer entities were included in the study. Only projects where at least 20 Tweets were found were included. Unfortunately, this systematic strategy only yielded three development projects. To ensure a larger sample size, we proceeded to use both Google and Twitter search functions to identify additional such projects, ultimately finding 21 development projects that met the pre-established criteria (Table 2).

All tweets used were scraped on May 15, 2020 through the open sourcemodule "GetOldTweets3". At first, we attempted to use the official Twitter module "tweepy" but due to restrictions on the tweets which could be sourced using the module with respect to age and quantity, we turned to "GetOldTweets3" which scrapes Twitter without utilizing an API key. The function we ran included the name (and other related nicknames[1] or aliases for the project) in conjunction with five descriptors, like this:

- "'ABC' development"
- "'ABC' redevelopment"
- "'ABC' project"
- "'ABC' construction"
- "'ABC' approval"

Once the tweets were gathered, the next step was determining whether each tweet was published by a bot. The initial plan was to use an already existing machine learning model to classify tweets, so the first attempt was using the Botometer, created by researchers at the University of Indiana. However, the scores output by the Botometer were unreasonably low, so this method was discarded.

Using the same test datasets used to construct the Botometer, we created our own machine learning model using the random forest classification technique in order to classify Twitter accounts as bots. The random forest classification approach was chosen due to its flexible, non-parametric nature and its ability to perform at a high enough level while being a convenient model for visualization.

[1] The names of some of the development projects were so generic that for several cases additional keywords were included for the sake of additional clarity. For example, Bulfinch Crossing is also known as 'One Congress' but One Congress is a very vague search term by itself so 'Boston' was added such that the query would be: "'One Congress' Boston project/development/etc."

We utilized the following five datasets to train our model, drawing on sets of both verified human Twitter accounts as well as self-identified bots, fake follower accounts, and celebrity human accounts (the test datasets are all publicly available at the Indiana University website.[2]):

1. "Verified-2019". Verified human accounts. Labels and user objects (Yang, et al. 2019a).
2. "Botwiki-2019". Self-identified bots from https://botwiki.org. Labels and user objects (Yang, et al. 2019a).
3. "Cresci-rtbust-2019". Manually annotated bot and human accounts. Labels and user objects (Mazza, et al. 2019).
4. "Vendor-purchased-2019". Fake follower accounts purchased from several companies. Labels and user objects (Yang, et al. 2019b).
5. "Celebrity-2019". Celebrity accounts collected as authentic users. Labels and user objects (Yang, et al. 2019b)

After calibrating the random forest model, we ran the Tweets we had collected for the 21 development projects through the model. For any Twitter user account that was classified as a bot, we monitored their activity and the extent to which their posts, likes, and re-posts appear to be shaping real world conversations. Lastly, we analyze a series of social media accounts to project the possible influence each bot may generate.

Results

We utilized the official Twitter API through "tweepy" to collect the user data corresponding to the tweets (scraped with GetOldTweets3) identified as being related to the 21 real estate development projects. We utilized "tweepy" instead of GetOldTweets3 to get the user information (our features) because there were no restrictions with respect to the official API and collecting user

[2] https://botometer.iuni.iu.edu/bot-repository/datasets.html

data (as opposed to tweets, which needed GetOldTweets3). A total of 41,191 Tweets were collected concerning 21 real estate development projects, with a median of 560 Tweets per project (Table 2). The dates of the Tweets ranged from May 9th, 2007 to May 15th, 2020. The Hudson Yards development had the most Tweets with 14,995 and the Millennium Place development had the least with only 19 related Tweets.

Table 2: Details of 21 development projects

Project	**City**	**Start Year**	**Development Type**	**Stage**	**Size (acres)**	**Tweets**
Perimeter Center East	Atlanta GA	2018	Mixed use	Concept	20	48
Centennial Yards	Atlanta GA	2019	Mixed use	Planning	50	1255
Port Covington	Baltimore MD	2014	Mixed use	In progress	235	2708
Millennium Place	Boston MA	2012	Luxury condo	Completed	N/A	19
Allston Yards	Boston MA	2019	Mixed use	Planning	11	143
One Congress	Boston MA	2019	Mixed use	Starting	23	202
“La Central: Bronx	Bronx, NY	2018	Mixed use	Phase 2 construction	27.5	248
Pacific Park	Brooklyn NY	2008	Mixed use	In progress	22	5825
Lincoln Yards	Chicago IL	2016	Mixed use	Planning	50+	2002
Commerce Point	Commerce GA	2019	Mixed use	Proposed	100	53
Belmont Park Arena	Elmont NY	2017	Arena, retail, train station	In progress	23	1168
SoFi Stadium	Inglewood CA	2015	Stadium	Construction	298	792
Miami World Center	Miami FA	2019	Mixed use	Construction	25	316
Riverside South	New York NY	1994	Mixed use	In progress	57	320
Hudson Yards	New York NY	2012	Mixed use	In progress	27	14995
Washington Terrace	Raleigh NC	2015	Mixed residential	Complete	23	81
Transbay	San Francisco CA	2005	Mixed use	In progress	71	5183
Candlestick Point	San Francisco CA	2008	Mixed use	Construction	280	2083

Treasure Island	San Francisco CA	2011	Mixed use	Construction	404	3204
Parkmerced	San Francisco CA	2011	Redevelopment	Planning	152	440
Capitol Crossing	Washington DC	2014	Redevelopment	In progress	N/A	559
Burnham Place	Washington DC	2015	Mixed use	Planning	69	44

We created a random forest classification model to classify twitter accounts either as humans or bots. Variables (or features) in the model were chosen through a review of the literature on bot identification and through a 5-fold cross-validation of 100 iterations (Table 3). For each tweet were attempting to classify as from a bot or not, we took the user data scraped with "tweepy" as our independent variable. For more detail, see the variable importance Figures 3 and 4 in the Appendix. Training-test dataset split was 80% - 20%. In total, there were 25,447 users in the datasets.

Table 3: Variable names and details.

Variables (features)	Description
Followers count	The number of people who follow the user.
Listed count	The number of times the user has been placed into a Twitter List. A List is a curated group of Twitter Users which can be created and followed by anyone.
Ratio of followers and following	Ratio of followers to following for the user.
Friends count	The number of people the user follows ("Following").
Statuses count	The number of original Tweets posted by the user
Default profile	A boolean value representing whether the user ever updated their profile.
Verified	Twitter has verified this account is the person indicated by the account name.
Status ratio	The ratio of original Tweets by the user and the number of Tweets they have favorited, chosen to see how much the content the user produces vs how much they consume.
Favorites count	The number of Tweets favorited by the user.
Default profile image	Boolean value representing whether the user ever updated their profile image.
Geolocation enabled	Boolean value representing whether the user enabled the geolocation feature, allowing others to detect their latitude and longitude (deprecated).
Translator enabled	Boolean value representing whether the user enabled the translation feature for Tweets (deprecated).

Is translator	Boolean value representing whether the user is a part of the Twitter translator community (deprecated).
Contributors enabled	Boolean value representing whether the user (deprecated).
Protected	Boolean value representing whether the user's tweets are considered protected and not public.

When using all the training datasets, the percentage of Tweets classified as coming from bots was 8.4% - a particularly low percentage given Varol et al.'s (2017) research that up to 15% of all Tweets are generated by bots. Therefore, we utilized Synthetic Minority Oversampling Technique (SMOTE) to create a training set such that 13.72% of the users we analysed were bots. SMOTE is a resampling method which helps deal with the issue of imbalanced datasets in classification problems. Traditionally, undersampling the majority class of variables has been used to increase the sensitivity of models. SMOTE resamples by oversampling the minority class while also under-sampling the majority class. This method has been found to be more than simply undersampling the majority class (Chawal, et al. 2002). For the SMOTE model, we adopted Varol et al.'s (2017) estimate that 15% of all Twitter users are bots to set our sampling settings for our test data, essentially pre-programming the model to expect that roughly 15% of users will be bots.

Two models were created, one from the SMOTE datasets and one from the non-SMOTE dataset, where no rebalancing settings were put in place and no pre-conceived percentage of users were expected to be bots (see Tables 4 and 5). The tables display each of the models' performance on a held-out test set which was 20% of the full 25,362 example data set, meaning that the models were evaluated on a dataset of 5072 examples.

Table 4: Model classification results on held-out test set (with SMOTE datasets).

	Precision	**Recall**	**f1-score**	**Support**
Bot	0.85	0.77	0.81	410
Human	0.98	0.99	0.98	4680
Accuracy			0.97	5090
Macro average	0.91	0.88	0.97	5090

Weighted average	0.97	0.97	0.97	5090
Train score	0.97			
Test score	0.97			

Table 5: Model classification results on held-out test set(with non-SMOTE datasets).

	Precision	**Recall**	**f1-score**	**Support**
Bot	0.91	0.68	0.78	410
Human	0.97	0.99	0.98	4680
Accuracy			0.97	5090
Macro average	0.94	0.84	0.88	5090
Weighted average	0.97	0.97	0.97	5090
Train score	0.98			
Test score	0.97			

Notes: Precision for a class represents the fraction of users correctly classified as a part of the class from all users identified as a member of the class from the model. Recall for a class represents the fraction of all users from the class identified by the model correctly.

The most interesting performance statistics we found in the models was the precision and recall for each model with regards to bot detection: the SMOTE based model has lower precision but has greater recall. This means that although the SMOTE based model may not select the bots as accurately from the pool of users, it picks up a larger proportion of the bots in general (see Tables 4 and 5). In terms of accuracy, the models perform approximately the same (0.97). Comparing F1-scores, the harmonic means of the precision and recall, also known as the Dice similarity coefficient (DSC), the SMOTE based model performs better with a F1-score for bot identification of 0.81 while the non-SMOTE based model has a F1-score of 0.78 (see Tables 4 and 5). Accuracy tends to depend more upon the number of True Positive and True Negative classification cases while the F1-Score depends more upon the number of False Positives and False Negatives. In the case of heavily imbalanced class distributions, the F1-score is a better measure of the performance of the model. Thus, it seems that the SMOTE based model performs better overall.

Using Model to Classify Tweets as Human or Bot

Figures 1 and 2 depict the results of bot percentages calculated for all developments as well as total bot counts.

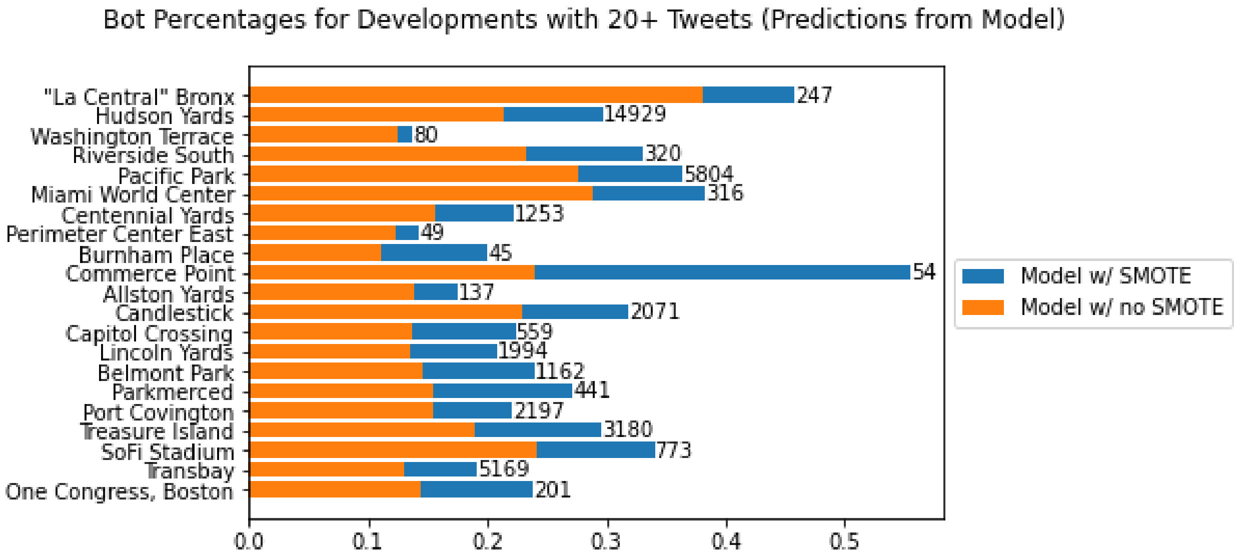


Figure 1: Relative number of bots identified for 21 Real estate development projects (with both SMOTE and non-SMOTE datasets).

The SMOTE model found the Commerce Point development project as having the highest proportion of bot-generated Tweets (55%) followed by “La Central” Bronx (43%). The SMOTE model found Tweets from all projects to be at least 15% bots. The non-SMOTE model detected at least 10% bots in all development projects, with “La Central” Bronx with the highest percentage, nearly 40%.

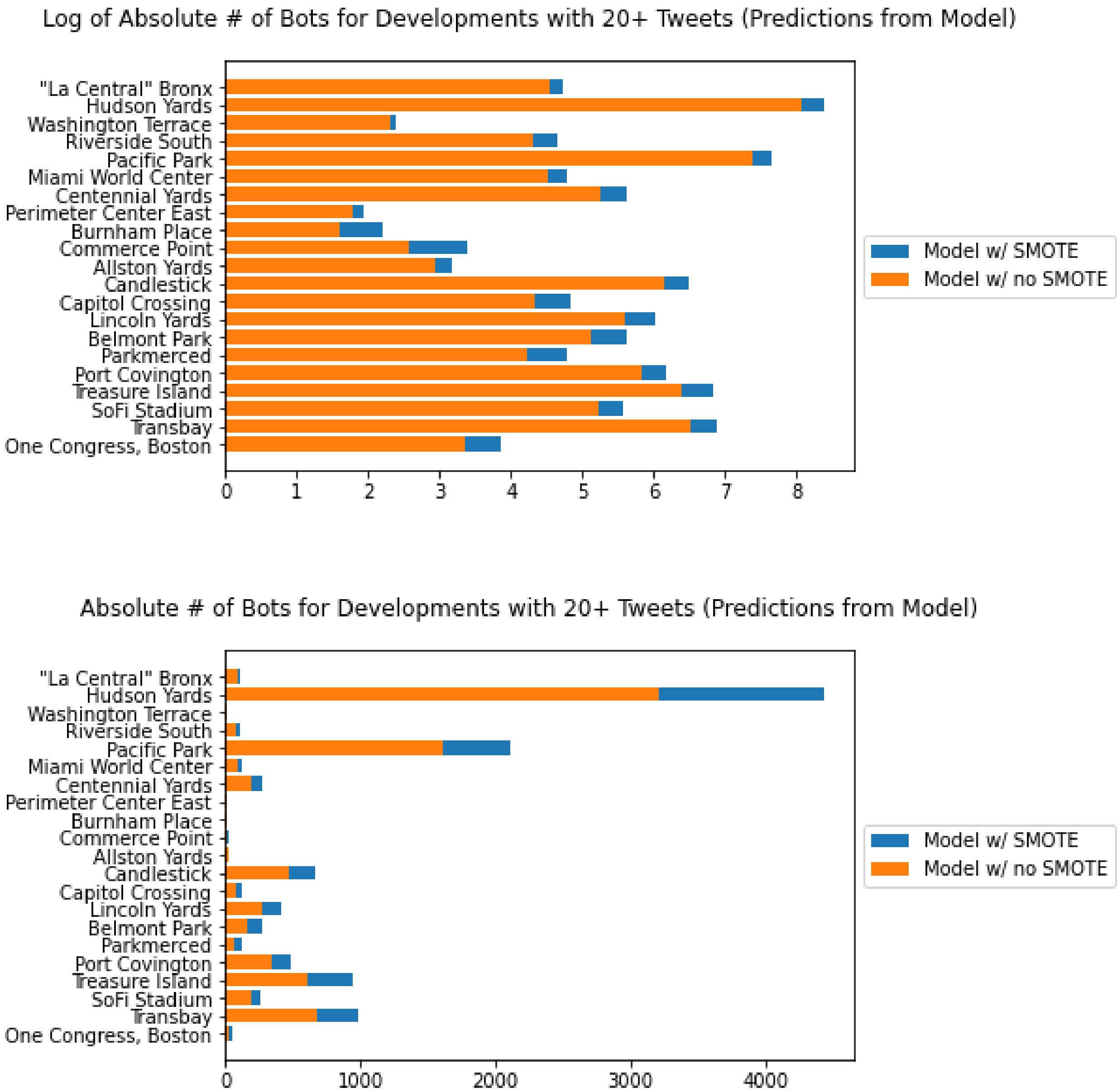


Figure 2: Log and Absolute number of bots identified for 21 Real estate development projects (with both SMOTE and non-SMOTE datasets).

Qualitative Review of Bot Activity and Impact

We randomly selected twenty bot-generated tweets from each development using a random number generator to qualitatively assess the account user, as well as the content of their tweets and their interaction and impact on Twitter. Of the sample of 420 tweets, 42 tweets were not accessible: 35 had been deleted since the data scrape, two were from private accounts, and five had been tweeted from now suspended accounts. Of the remaining 378 tweets, 368 came from different accounts. Our manual reading of the tweets found that 72% (272) were clearly from bot or automated accounts. Of the remaining 126 tweets authored by what appeared to be real users (but could be sophisticated bots), 33 came from news reporters or communications/public relations professionals, and another 17 came from realtors or professionals related to the real estate industry.

In fact, news and real estate were the most prominent themes throughout the entire 378 tweet sample. These two categories of user account for almost half of all the tweets in the sample: 94 tweets came from users whose profile description clearly associated them with a news outlet, and 72 were from real estate related accounts. In terms of political or ideological positionality, the majority of the user descriptions were neutral. Only 30 accounts had positions clearly outlined in their user descriptions: 25 were advocates of a particular planning-related theme such as heritage, transit, smart growth, or racial equity, while five accounts focused on inflammatory messaging such as conspiracy theories, strong anti-development sentiments, and even seemingly random offensive tweets.

Regarding engagement and activity, the 378 tweets had very little interaction with other users. 262 of them had zero likes, retweets, or comments. The tweet with the most interaction had 1224 likes, 150 retweets, and 41 comments, however, it was very loosely related to theme of our examination. Although it was technically about the SoFi development, the content of the tweet focused on COVID-19 regulations. In fact, five of the 17 tweets that received ten or more likes relate to COVID-19 regulations on various development sites. Nine of the 17 tweets are from news bots or reporters.

Looking past the user descriptions and engagement to concentrate on the content of the sample of 378 tweets, one major trend becomes clear: outside links. Almost 80% of the tweets (302) contained a link to an outside article or website. 85 of the links were dead, 19 were not accessible due to a paywall, and 10 brought up completely unrelated content. Of the remaining 188 links that were relevant to the developments, 45 were openly supportive of the developments and 26 were critical. Critical items included a *Curbed New York* article describing the “state of Brooklyn’s Pacific Park Megaproject” which highlighted that nearby residents “will get to experience the joys of living next to a massive construction site for years” and an article in the *Miami Herald* which examined potential inappropriate financial dealings between the developer of the Miami World Center and local elected officials. Supportive items included a *4NBC Washington* story touting that the Capital Crossing project will improve traffic flow, a *PoliticalDog101* blog describing the extensive support for the Belmont Park development, and a *Baltimore Business Journal* article praising the size of the Port Covington development. It is interesting to note that half of the tweets with links critical of the developments were posted by what appear to be real users, while less than a quarter of the tweets with supportive links were authored by what appear to be real users. Perhaps most telling is that over half of the relevant links were neutral. This finding reflects the prevalence of news-bots in the account analysis. A lot of the automated online discussion on the developments appears to simply be generic news updates.

Conclusion

Using machine learning technique, we built two Random Forest Models which performed at a high degree of both accuracy and precision in predicting whether a Tweet was produced by a bot or not. We then used our sample of Tweets from 21 real estate development projects to assess how prevalent bots were within this domain of social media. We found that at least 10% and in one case 50% of Tweets in some of these development projects were being generated by bots. While this study was limited and our findings varied across SMOTE and non-SMOTE models, these finding suggest that the use of bots in real estate development communications online may be widespread and potentially harmful.

There are some important limitations to this research. First, by scraping the Tweets we avoided the established Twitter system for obtaining data, raising legal and ethical issues. Scraping has been an increasingly common modality for scholars in studying social media activities and offers insights on the human and urban condition otherwise impossible to capture (Hollander, et al. 2016). Also, some of the developments we studied had been finished for some time meaning many tweets/users relevant to one of these older developments who were bots could have been removed by Twitter's own bot moderators. The bot detection approach we applied is not nearly as accurate as human detection so there is room for error in the bot numbers we identified. The random forests classifier was built upon data taken from botometer; it was imbalanced as there were around 20,000 more humans than bots including at least 10,000 verified accounts (humans with the blue check next to handle). These verified accounts are not as frequent as they seemed in the dataset we used to train, but they were needed to create a more natural ratio of humans to bots.

To address some of these weaknesses, future investigations could employ more elaborate discourse analysis to better understand bot language and communication activity. Such investigations could track the conversations online, monitor them over time, and develop an analytical framework to assess the likelihood that a user is a social bot or not. Ultimately, the hope would be that scholars could construct social network maps to assess how influential each follower is, so that we may project the possible influence (and harm to community well-being) each bot could potentially generate. Regarding the ethical dimensions of scraping, future research could seek out more formal research collaborations with Twitter or other social media platforms to partner on follow-up studies.

This work lays the foundation for additional research involving the development of custom-designed Twitter bots to test a range of cyber influencing strategies. Each bot could have a core set of artificial intelligence: the ability to follow likeminded users, to retweet and "like" Tweets on a predetermined set of topics, and to re-post Reddit images with short phrases employing the Microsoft Computer Vision software. This type of experimental research would allow researchers to construct social network maps to assess influential and to project the possible influence each bot may generate, painting a fuller picture of the extent to which the real estate development community is vulnerable to disruption online.

References: Abokhodair, N., Yoo, D., & McDonald, D. W. (2015). *Dissecting a social botnet: Growth, content and influence in Twitter.* Paper presented at the Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing.

Afzalan, N., & Evans-Cowley, J. (2015). Planning and social media: Facebook for planning at the neighbourhood scale. *Planning Practice & Research, 30*(3), 270-285.

Alarifi, A., Alsaleh, M., & Al-Salman, A. (2016). Twitter turing test: Identifying social machines. *Information Sciences, 372*, 332-346.

Arribas-Bel, D., Kourtit, K., Nijkamp, P., & Steenbruggen, J. (2015). Cyber cities: Social media as a tool for understanding cities. *Applied Spatial Analysis and Policy, 8*(3), 231-247.

Badawy, A., Ferrara, E., & Lerman, K. (2018). *Analyzing the digital traces of political manipulation: the 2016 Russian interference Twitter campaign.* Paper presented at the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).

Bastos, M., & Mercea, D. (2018). The public accountability of social platforms: lessons from a study on bots and trolls in the Brexit campaign. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376*(2128), 20180003.

Bessi, A., & Ferrara, E. (2016). Social bots distort the 2016 US Presidential election online discussion. *First Monday, 21*(11-7).

Boshmaf, Y., Muslukhov, I., Beznosov, K., & Ripeanu, M. (2011). *The socialbot network: when bots socialize for fame and money.* Paper presented at the Proceedings of the 27th annual computer security applications conference.

Chmielewski, S., Samulowska, M., Lupa, M., Lee, D. J., & Zagajewski, B. (2018). Citizen science and WebGIS for outdoor advertisement visual pollution assessment. *Comput. Environ. Urban Syst., 67*, 97-109.

Edwards, C., Edwards, A., Spence, P. R., & Shelton, A. K. (2014). 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.

Ertiö, T.-P., & Bhagwatwar, A. (2017). Citizens as planners: Harnessing information and values from the bottom-up. *International Journal of Information Management, 37*(3), 111-113.

Evans-Cowley, J. S. (2010). Planning in the age of Facebook: the role of social networking in planning processes. *GeoJournal, 75*(5), 407-420.

Evans-Cowley, J. S., & Conroy, M. (2006). The growth of e-government in municipal planning. *Journal of Urban Technology, 13*(1), 81-107. doi: 10.1080/10630730600752892

Evans-Cowley, J. S., & Griffin, G. P. (2011). Micro-participation: The role of microblogging in planning. *Available at SSRN 1760522*.

Evans-Cowley, J. S., & Hollander, J. (2010). The new generation of public participation: Internet-based participation tools. *Planning Practice & Research, 25*(3), 397-408.

Ferrara, E. (2018). Measuring social spam and the effect of bots on information diffusion in social media *Complex Spreading Phenomena in Social Systems* (pp. 229-255): Springer.

Firmstone, J., & Coleman, S. (2015). Public engagement in local government: the voice and influence of citizens in online communicative spaces *Information, Communication & Society, 18*(6), 680-695. doi: 10.1080/1369118X.2014.986180

Fredericks, J., & Foth, M. (2013). Augmenting public participation: enhancing planning outcomes through the use of social media and web 2.0. *Australian planner, 50*(3), 244-256.

Hanzl, M. (2007). Information technology as a tool for public participation in urban planning: a review of experiments and potentials. *Design Studies, 28*(3), 289-307.

Hasanzadeh, K., Kajosaari, A., Häggman, D., & Kyttä, M. (2020). A context sensitive approach to anonymizing public participation GIS data: From development to the assessment of anonymization effects on data quality. *Computers, Environment and Urban Systems, 83*, 101513.

Healey, P. (2003). Collaborative Planning in Perspective. *Planning Theory, 2*(2), 101-123. doi: 10.1177/14730952030022002

Hegelich, S., & Janetzko, D. (2016). *Are social bots on Twitter political actors? Empirical evidence from a Ukrainian social botnet.* Paper presented at the Tenth International AAAI Conference on Web and Social Media.

Horelli, L., Saad-Sulonen, J., Wallin, S., & Botero, A. (2015). When self-organization intersects with urban Planning: Two cases from Helsinki. *Planning Practice & Research, 30*(3), 286-302.

Innes, J. E. (1995). Planning Theory's Emerging Paradigm: Communicative Action and Interactive Practice. *Journal of Planning Education and Research, 14*(3), 183-189. doi: 10.1177/0739456X9501400307

Klosterman, R. (1997). Planning support systems: a new perspective on computer-aided planning. *Journal of Planning education and research, 17*(1), 45-54.

Klosterman, R., & Landis, J. D. (1988). Microcomputers in US Planning: Past Present and Future. *Environment and Planning B: Planning and Design, 15*(3), 355-367. doi: 10.1068/b150355

Lin, Y., & Geertman, S. (2019). *Can Social Media Play a Role in Urban Planning? A Literature Review.* Paper presented at the International Conference on Computers in Urban Planning and Urban Management.

Mandarano, L., Meenar, M., & Steins, C. (2010). Building social capital in the digital age of civic engagement. *Journal of planning literature, 25*(2), 123-135.

Paquet-Clouston, M., Bilodeau, O., & Décary-Hétu, D. (2017). *Can We Trust Social Media Data? Social Network Manipulation by an IoT Botnet.* Paper presented at the Proceedings of the 8th International Conference on Social Media & Society.

Ratkiewicz, J., Conover, M. D., Meiss, M., Gonçalves, B., Flammini, A., & Menczer, F. M. (2011). *Detecting and tracking political abuse in social media.* Paper presented at the Fifth international AAAI conference on weblogs and social media.

Shao, C., Ciampaglia, G. L., Varol, O., Yang, K.-C., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature communications, 9*(1), 4787.

Stieglitz, S., Brachten, F., Berthelé, D., Schlaus, M., Venetopoulou, C., & Veutgen, D. (2017). *Do social bots (still) act different to humans?–Comparing metrics of social bots with those of* *humans.* Paper presented at the International Conference on Social Computing and Social Media.

Subrahmanian, V., Azaria, A., Durst, S., Kagan, V., Galstyan, A., Lerman, K., . . . Menczer, F. (2016). The DARPA Twitter bot challenge. *Computer, 49*(6), 38-46.

Wald, R., Khoshgoftaar, T. M., Napolitano, A., & Sumner, C. (2013). *Predicting susceptibility to social bots on twitter.* Paper presented at the 2013 IEEE 14th International Conference on Information Reuse & Integration (IRI).

Williamson, W., & Parolin, B. (2012). Review of Web-Based Communications for Town Planning in Local Government. *Journal of Urban Technology, 19*(1), 43-63. doi: 10.1080/10630732.2012.626702

Williamson, W., & Parolin, B. (2013). Web 2.0 and social media growth in planning practice: A longitudinal study. *Planning Practice and Research, 28*(5), 544-562.

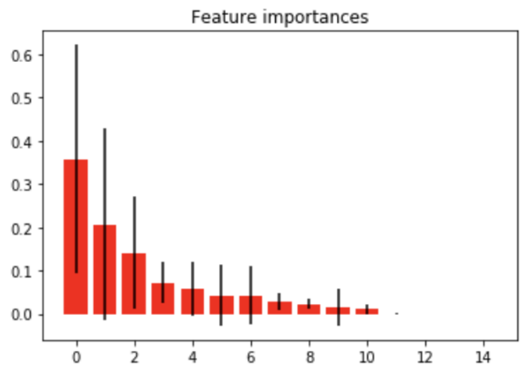
Williamson, W., & Ruming, K. (2015). Assessing the effectiveness of online community opposition to precinct planning. *Australian Planner, 52*(1), 51-59.

Williamson, W., & Ruming, K. (2020). Can social media support large scale public participation in urban planning? The case of the# MySydney digital engagement campaign. *International Planning Studies, 25*(4), 355-371.

Wilson, A., Tewdwr-Jones, M., & Comber, R. (2017). Urban planning, public participation and digital technology: App development as a method of generating citizen involvement in local planning processes. *Environment and Planning B: Urban Analytics and City Science*, 2399808317712515. doi: 10.1177/2399808317712515

Woolley, S. C., & Howard, P. N. (2018). *Computational propaganda: political parties, politicians, and political manipulation on social media*: Oxford University Press.

```
Feature ranking:
0. followers_count (0.358352)
1. listed_count (0.207207)
2. ratio (0.141399)
3. friends_count (0.071969)
4. statuses_count (0.057783)
5. default_profile (0.043120)
6. verified (0.042691)
7. status ratio (0.027645)
8. favourites_count (0.022360)
9. default_profile_image (0.015988)
10. geo_enabled (0.011210)
11. is_translation_enabled (0.000278)
12. is_translator (0.000000)
13. contributors_enabled (0.000000)
14. protected (0.000000)
```




```
{'max_depth': 8,
 'max_features': 'auto',
 'max_leaf_nodes': 268,
 'min_samples_leaf': 4,
 'min_samples_split': 6,
 'n_estimators': 33}
```

Figure 4: Feature/Variable importance (with non-SMOTE datasets).