

EXECUTIVE SUMMARY OF FINAL REPORT TO
Lambda Alpha International's Land Economics Foundation (LEF)



USING TWITTER BOTS
TO ASSESS RISKS TO LAND
DEVELOPMENT AND PLANNING

By:
Justin B. Hollander, PhD, FAICP (Tufts University)
Maxwell Hartt, PhD (Queens University)
Ruth Potts, PhD (Cardiff University)

With contributions from:
Alexander Seto
Minyu Situ
Anna Yuen
Rachel Herman

Using Twitter Bots to Assess Risks to Land Development and Planning

Final Report to Lambda Alpha International's Land Economics Foundation

(LEF)

Executive Summary

Justin B. Hollander, PhD, FAICP (Tufts University)

Maxwell Hartt, PhD (Queen's University)

Ruth Potts, PhD (Cardiff University)

With contributions from: Alexander Seto, Minyu Situ, Anna Yuen, and Rachel

Herman

August 25, 2020

This report examines the role and potential risk of automated social media accounts in the participatory planning process through an empirical investigation of real estate development projects in the U.S., summarizing research conducted by Tufts University, Cardiff University, and Queen's University with support from the Lambda Alpha International's Land Economics Foundation.

The rapid growth and massive uptake of social media has resulted in a surge in public interest to engage with others around key land development topics. Due to the low cost and

high potential engagement, planners and policymakers have been quick to open electronic channels of participation to inform the decision-making process. Doing so has created an opportunity for subversion from groups with alternate and possibly nefarious interests. Anecdotally, we have found that automated social media accounts have been used to further inflate the voice, and therefore influence, of subversive groups in the land development and planning process. While scholars have begun to examine how tech-savvy social media users are manipulating political discourse through the medium of Facebook and Twitter, no research has yet sought to examine the potential harm that such manipulation can and does cause at the local level in a planning context.

This project explored the risks that social media manipulation might pose to online community discourse around land development and planning topics. To begin to gauge this risk, we reviewed the theoretical and empirical literature on the topic, conducted an investigation of bot activity on Twitter surrounding 21 major real estate projects in the U.S. and developed recommendation for additional research and policy implications.

Employing a convenience sampling strategy, we first culled major urban planning and real estate development professional journals, magazines, and websites for the names and locations of large-scale real estate development projects being proposed or developed over the last three years 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.

A total of 41,191 Tweets were collected concerning 21 real estate development projects, with a median of 560 Tweets per project. Once the tweets were gathered, the next step was determining whether or not each tweet was put out by a bot. We created our own machine learning model using the random forest classification technique in order to classify twitter accounts as bots. We utilized the four 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. After calibrating the random forest model (using both weighted, SMOTE [Synthetic Minority Oversampling Technique], and unweighted, non-SMOTE varieties), 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 analyzed a series of social media accounts to project the possible influence each bot may generate.

Bot percentages were calculated for all developments as well as total bot counts:

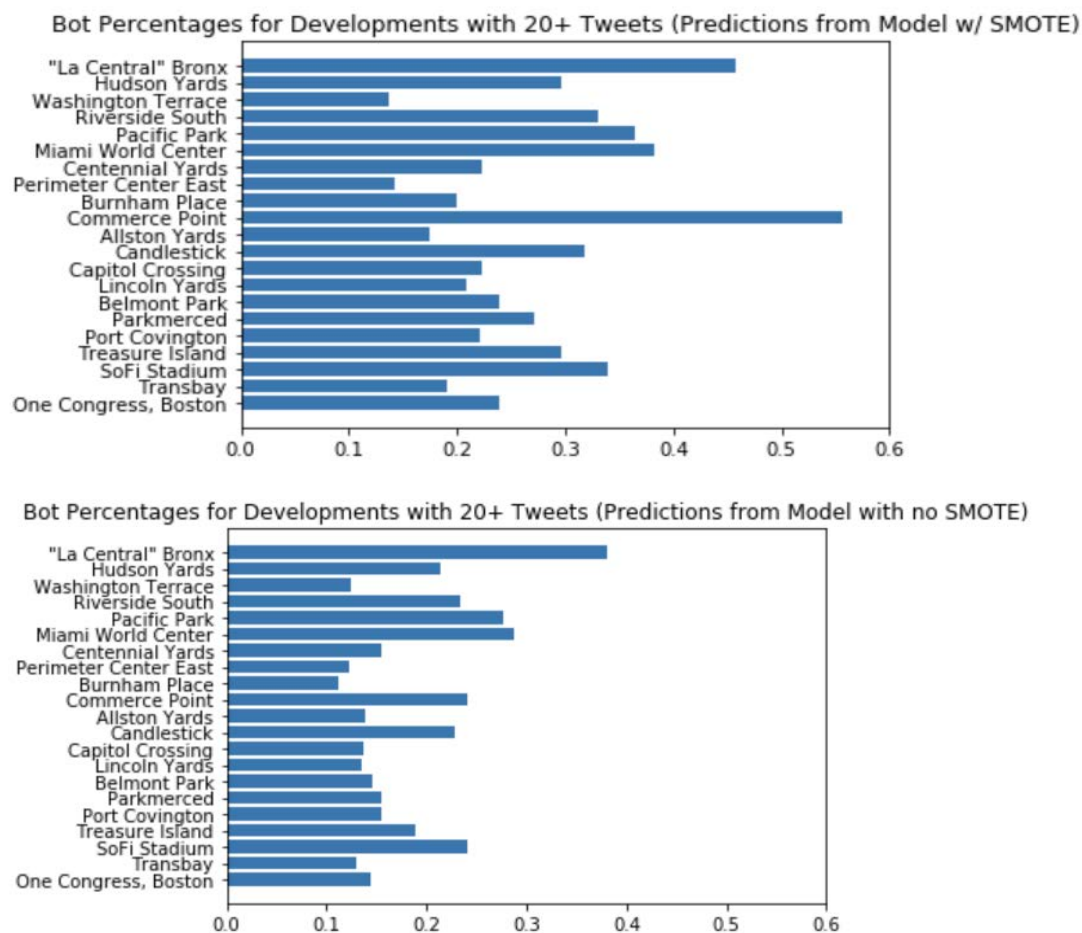


Figure 1: Relative number of bots identified for 21 Real estate development projects (with both weighted and un-weighted datasets)

Using an advanced 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 upwards of 50% of Tweets in some of these development projects were being generated by bots. Among the Tweets in our sample, it was the Commerce Point development project that was characterized by the SMOTE model as have the highest percentage of Tweets, with 55%, followed by “La Central” Bronx with 43%. Using the SMOTE model, all the projects had at least 15% bots, but the non-SMOTE model showed less variation detecting bots in at least 10% of all development projects, with “La Central” Bronx with the highest percentage, nearly 40%. These findings suggest that the use of bots in real estate development communications online is widespread and potentially harmful.



Figure 2: Tweets/users identified as bots by both models

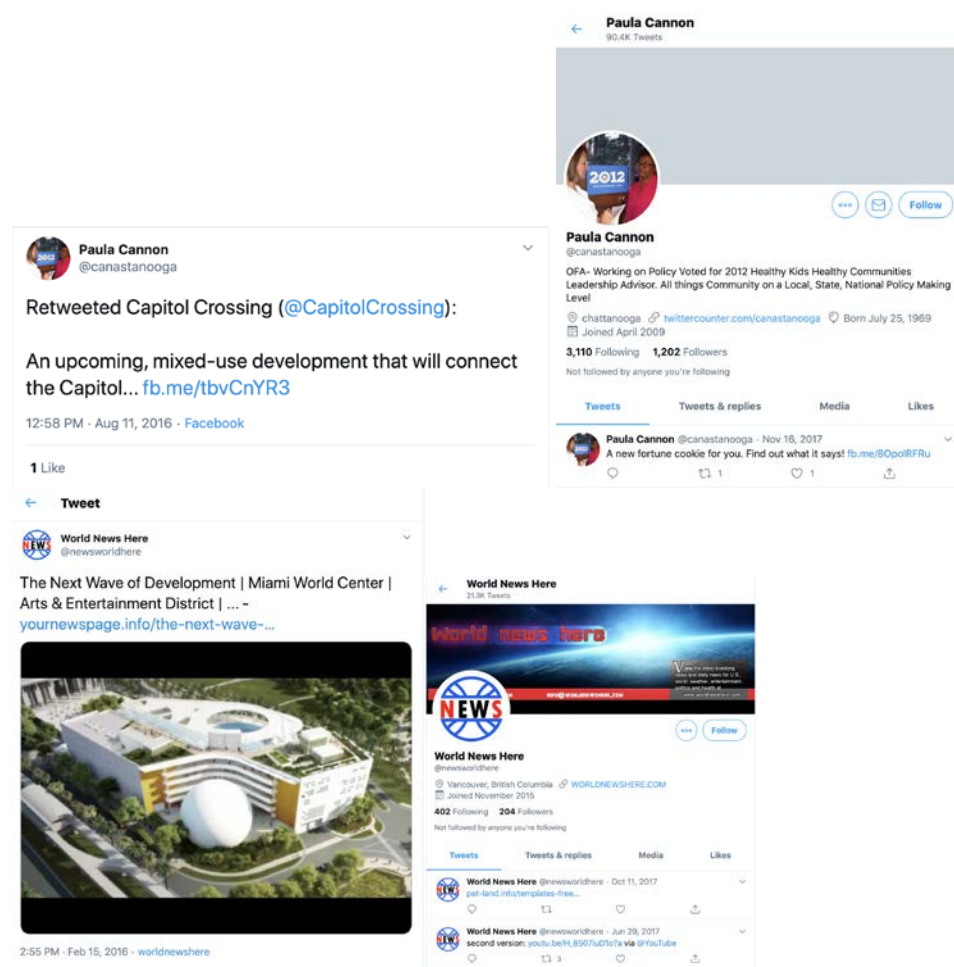


Figure 3: Additional Tweets/users identified as bots by both models

Due to the sheer number of bots identified by our machine learning model, we randomly selected one bot-generated tweet 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, see examples in Figure 2 and 3. It is interesting to note that in our qualitative review of tweets by bots, they had very little interaction with other users. 14 of them had zero likes, retweets or comments. Of the subsample, the tweet with the most interaction had 4 likes and 5 retweets.

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 how influential each follower is 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.