Visualizing Tweets from Confirmed Fake Russian Accounts

Stephen Hsu, David Kes, and Alark Joshi, University of San Francisco, San Francisco, CA, USA

Abstract

Social media has played a huge role in the 2016 US Presidential Elections. In this paper, we present the results of an exploratory visualization of 200,000+ tweets from confirmed fake twitter accounts. We analyze the user accounts by examining their user names, descriptions/bios, tweets, tweet frequency, and content. We found that they made themselves relatable using political and religious beliefs and then used their influence by joining into popular hashtags on Twitter and posting strongly polarizing tweets at crucial times such as debates and primaries in the election cycle.

Introduction

Social media plays an important role in today's world with its influence on entertainment, sports, news, and even elections [11]. With the fervor of the Presidential election being skewed by Russian interference on Twitter and Facebook as well as the notorious Facebook / Cambridge Analytica scandal [21], it was obvious that there was some interference and we wanted to explore the data. We wanted to explore who these "fake users/twitter bots" were and their characteristics. Specifically, we asked questions such as "Who are these fake users disguising as?" "How are these fake users influencing individuals people?"

We applied a combination of natural language processing techniques as well as exploratory data visualization techniques to examine the fake Russian users Tweet data. In particular, we analyzed information about the users and their tweets through the following questions:

1. Who these fake users pretend to be - names and descriptions
2. When the fake accounts were being created
3. When the fake accounts were most active in tweeting
4. What topics the accounts were covering

We also dug into the data and present the results of a case study of the most successful users. Specifically, we examined the velocity, polarity, and subjectivity of the tweets from these influential accounts. We found that the top twenty influential accounts posted polarizing tweets especially on weekends when scandals/announcements occurred. Later these accounts changed their approach by joining trending hashtags on twitter and injecting propaganda into those tweets.

Related Work

Social media is widely being used to analyze current sentiments, events, movement patterns, and so on. [3, 5]. With the large amount of data available in terms of text, images, videos, geolocation, hashtags, and so on, such analysis can be invaluable and provide insight into the data.

Visualizing Social Media Content

Diakopoulous et al. [7] created *Vox Civitas* - one of the early visual analytics interfaces that aggregated and annotated social media data with interactive visualizations that can help journalists summarize social media reactions for high visibility events such as the State of the Union. Abdullah et al. [1] introduced a way to measure happiness in a society by evaluating smiles through their *Smile Index* that analyzes smiles in 9 million geo-located tweets over a year. Hochman and Manovich [12] presented an innovative approach to visualize the photographs that were posted on social media. They identified networks, superimposed locations of photos on geographic locations, identified patterns from the images, and so on. Mallela et al. [14] introduced a system - CEST (City Event Summarization Tool) that is agnostic of events and data, but is able to capture sentiments and events dynamically as the data streams in. Miranda et al. [16] introduced a system - *Urban Pulse* that adopts techniques from computational topology to identify the signature of a city. They convert a variety of social media input to scalar functions that change over time and analyze it to identify the "pulse" of a city. Their approach is agnostic of a city and they demonstrate its application in New York and San Francisco. Xu et al. [24] developed a streamgraph-style interface to visualize the ebbs and flows of various topics being discussed in social media. They showcased their techniques on the 2012 United States Election and the Occupy Wall Street movement.

Visualizing Traffic and Movement

Endarnoto et al. [8] developed innovative ways to use twitter data to extract traffic conditions that can be consumed on a mobile device. Chen et al. [6] provide an approach to identify movement patterns using social media data. Their technique works on sparse data using geolocation and incorporate uncertainty visualization. This conveys the probability of the movement patterns using their heuristic rather than communicate sparse information with certainty.

Event and Outlier Detection

Xia et al. [23] introduced a system that allows users to see patterns and outliers in a city by analyzing social media data. They use geo-tagged photos in their system to analyze patterns and flag unusual activities. In subsequent work, Xia et al. [22] improved their approach to identify 'new' events in a city based on spatio-temporal analysis of social media data. They use a combination of Twitter and Instagram posts to validate their findings and to further find events that have low spatial and temporal deviation. Ferracani et al. [9] developed a web interface that allows the identification of local events in an urban environment using statistical methods. Giridhar et al. [10] developed a novel unsupervised approach to fuse multiple social media data (specifically Twitter and Instagram posts) and correlate events being detected across various social platforms. Borges et al. [3] provide an overview of

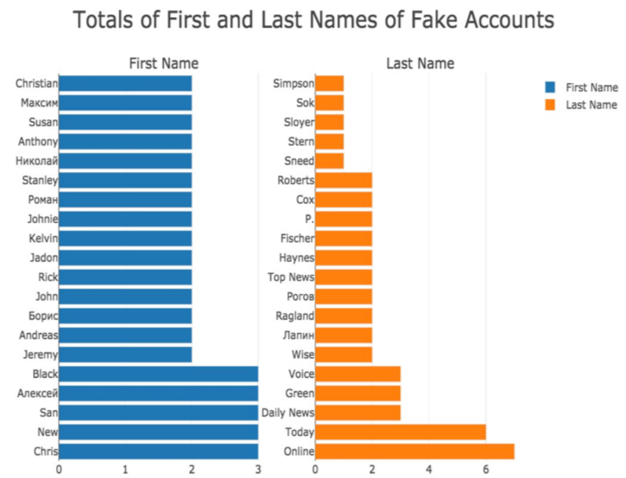

Figure 1. *First and Last "Names" of the Users. Common American Names, Respectable sounding news outlets, and Foreign names were the three major categories of all the users in the data.*

the various ways in which events can be detected in urban environments. They show how events such as music concerts too can be identified using social media. Chen et al. [5] provide a similar overview on the kinds of visual analytics that can be performed on social media data. They take a comprehensive approach to social media using a variety of social media such Twitter, Flickr, Foursquare, Sina Weibo, and so on.

Approach

To understand the influence of Twitter on elections, we started by looking into Twitter data that was linked to Russia during the 2016 U.S. Presidential Elections. We used the data published by NBC News [17] that contained more than 200,000 tweets that Twitter has tied to "malicious activity" from Russia-linked accounts to the election. These accounts coordinated to work together as part of a large network and sent out thousands of inflammatory tweets.

Our goal was to uncover any underlying structures in the data and detect any outliers or anomalies through the use of exploratory interactive visualization. We will first show the results of exploring the entire dataset to highlight characteristics of all the user accounts and then present a case study exploring the tweets of the top 20 most prolific and influential user accounts.

Who are these fake users?

To understand the user accounts of fake users, we first explore *just* the user names and found that there are three clear categories of the fake account names. The first combination consists of American sounding first names such as "Chris", "Rick" or "Jeremy" combined with American sounding last names such as "Green," "Roberts," or "Cox." The second combination consists of formal sounding news sources such as "Washington Online" or "Atlanta Today." Finally, the third combination consists of purely foreign names. From this, we can see that its often difficult to tell which accounts are fake based off the name alone as it could be any average Joe, news site, or just a foreign individual with an unrecognizable name. Figure 1 shows a histogram that conveys the distribution of the First and Last "Names" of all the users in the data.

What are their profile descriptions?

Once we had an understanding of the user account names, we turned our attention to the description/bio of the users. Every Twitter user has the option of including a short description of themselves and the description frequently helps a reader separate a user from a tweet bot. Users often post their summaries and ideologies for others to see. Specifically, Figure 2 shows one topic in common with the fake accounts: **religion**. By using words like "God," "InGodWeTrust," and "GodBlessAmerica," the fake accounts became relatable to a large group of people. Other bios include *black lives matters*, *Repeal Obamacare* and official news sounding descriptions (such as "sports," "weather," and "official") as well as foreign topics. Therefore, by quickly, relating to this fake user, users are more likely to follow or agree with the fake accounts tweets. These fake accounts used the well-known attitude similarity work by Byrne [4] that states that individuals with similar attitudes and beliefs are attracted to each other, whereas dissimilarity results in repulsion [19]. In 2014, Balmaceda [15] found that when interacting online individuals are attracted to others with similar opinions and beliefs.

Figure 2 shows a visualization of topics estimated using Latent Dirichlet Allocation (LDA) [2]. The figure shows a visualization of the LDA clusters of topics on the left and a histogram of the related terms for the user selected topic on the right. In the figure, the user selected topic cluster 7 and is shown a histogram of related terms that contain controversial terms related to black lives matter, Obama, police, AmericaFirst, and so on. Many of these topics were controversial during the election and were being debated and discussed by the politicians as well as users online. We used the LDAVis [20] tool to generate the figure.

When were they made?

The next important issue to consider is when these fake accounts were created. Figure 3 shows a timeline of when the accounts were created. Of the 454 accounts deemed to be fake Russian accounts, we can see the creation of the fake accounts started in 2009 and reached its peak of creations in 2013 before slowly lessening all the way to the start of 2017. Interestingly, this means the majority of fake accounts were created *years* before the actual 2016 Presidential election, perhaps to cause strife and division amongst US readers well in advance of the election.

Where are they from?

We then explored the origin of these accounts and discovered that of the 454 values, approximately half the values were *missing*. Of the 287 locations listed, 124 were listed as some form of "United States", 68 were listed as a large metropolitan cities in the United States (for example, San Francisco, New York, Atlanta, Los Angeles), and 37 values were in foreign countries, and the remaining 58 values were imaginary like "located at the corner of happy and healthy" or "the block down the street." Therefore, since the data was missing for a majority of the users, and most likely fake, we opted from analyzing the data further.

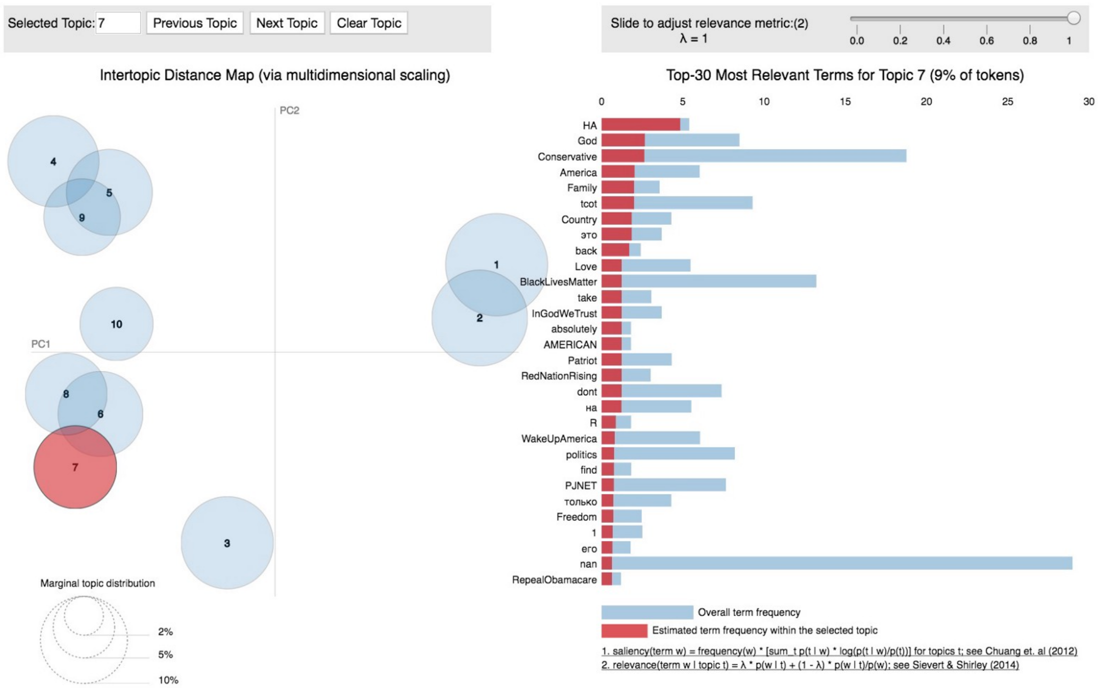

Figure 2. *Topics Within the Descriptions of Fake Users' Profiles. The left half of the figure shows the various topics that can be selected and the right half of the figure shows a histogram of the terms in that topic. In this figure, we picked Topic 7 on the left and related terms for that topic such as God, Conservative, Wake Up America, and so on are shown on the right. We used the LDAVis [20] tool to analyze the profile descriptions.*

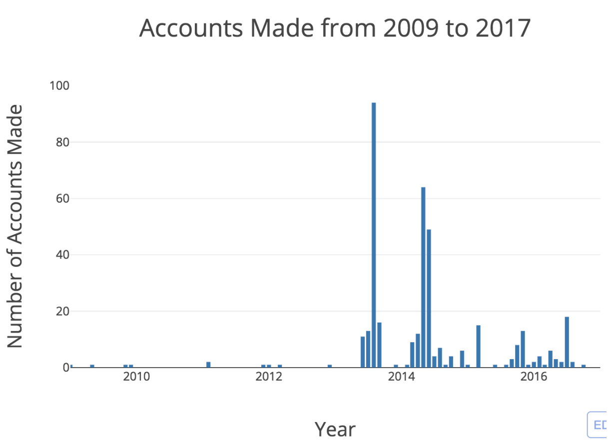

Figure 3. *Russian Fake Accounts Created from 2009 to 2017. Some accounts were created well in advance of the 2016 elections and were probably being explored for their influence on the 2014 midterm elections but were then kept alive with regular tweets to increase their influence.*

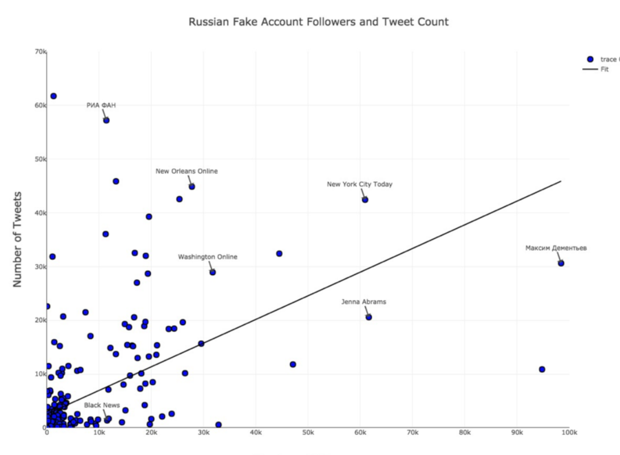

Figure 4. *Followers vs. Number of Tweets of Fake Accounts. As can be seen here, users who tweeted frequently had high numbers of followers. Outliers such as the fake Jenna Abrams account whose racist and controversial tweets were being discussed in the mainstream media can be seen here.*

How influential are these accounts?

We next wanted to explore the impact these fake accounts were having on the users of Twitter. We use the number of followers as a metric to see which accounts could be considered "influential." Figure 4 plots the number of tweets versus the number of followers and we can see the number of Tweets increasing with the number of followers. This makes sense as these fake accounts are leveraging their popularity on social media to reach

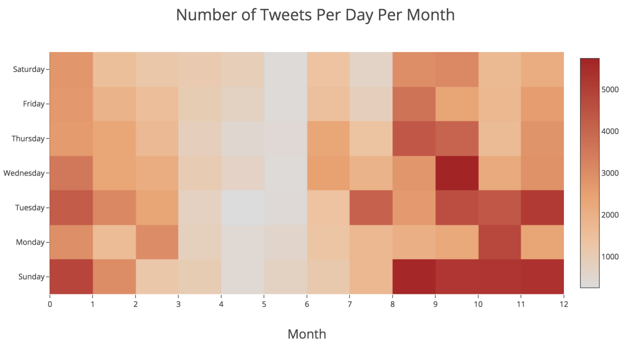

Figure 5. *Heat Map of Fake User Tweet Activity shows that Sundays and Tuesdays especially as the elections got closer (August-December) were more frequent for the fake accounts.*

out and influence more individuals. One such notable fake user is the infamous Jenna Abrams account, whose racist, controversial, and fake tweets were at one point covered in mainstream media. Given the large number of followers that these fake accounts had amassed, it does seem that these accounts were influential on the Twitter platform.

When are they posting?

We wanted to explore the days of the week when these fake accounts would post more frequently than others. We generated a heat map to visualize the days of the week and aggregated the data according to the month in the year. Figure 5 shows the above heat map where we can see that the fake users are predominantly posting on *Sundays* and *Tuesdays* in the later months of the year such as August, September, October, November, and December as the elections got closer. Based on the heatmap, we can determine that there is a pattern and that the tweets were not randomly posted. The fake users clearly understand the influence of their content and that it may be retweeted with more individuals on weekends rather than weekdays in the later half of the year when the election takes place.

What are they saying?

Figure 6 shows a screen shot of the Topical Modeling of the Tweet Content. Similar to the descriptions of the fake users, we examined the topics covered within the actual tweet content using LDA to cluster the data into topic clusters with terms. Figure 6 shows that the Black Lives Matters and other racial subject matters were one such topic that the Russian accounts targeted with words such as “police,” “blacklivesmatter,” “crime,” and references to the shooting in San Bernardino, particularly about the perpetrator being of minority descent. Other topics that we discovered were being propagated by the users consisted of Hillary Clintons private email server, ISIS, pro-Trump slogans, slanderings of the election debates, and school shootings. They added their polarizing opinions to these sensitive events and topics that were already being discussed actively on social media.

Case Study - The top 20 users

Now that we have explored the various characteristics of the fake account, we explore the data further in the form of a case study of the top 20 users. We used Figure 4 to determine the 20 most influential fake users to examine their tweeting behavior in terms of velocity, sentiment, and subjectivity over time of the top 20 followed fake Russian accounts.

Tweet Velocity

Figure 7 shows the overall volume of tweets from the 20 fake accounts. In terms of pure tweet volume, we can see a trend of the fake accounts being almost nonexistent until around June 2016, at which point the volume of Tweets increases dramatically: reaching its apex in October 2016. The tweets then tumble down in volume after November 2016 (election month), with one last resurgence around December 2016 before going back to an almost inactive state. This trend shows the opportunistic behavior of the fake accounts, tweeting at the most tense and vital points of the election fervor.

Tweet Sentiment and Subjectivity

Detecting the Sentiment and Subjectivity of a tweet was done using a Naïve Bayes classifier from the Textblob [13] package in Python. Textblob uses the NLTK Naïve Bayes classifier [18] for classification. If the word has never been seen before, the classifier ignores that word, otherwise, the classifier determines what the polarity or the subjectivity is for each specific word and uses the Bayes Rule to determine what the polarity or subjectivity is of the entire sentence.

The Sentiment could either be positive or negative and was scored on a +1 (most positive) to -1 (most negative) scale. Subjectivity was also computed as a value between 0 and 1, where 0 is very objective and 1 is very subjective. Subjective tweets are those where personal opinions and biases are clearly reflected, whereas objective tweets are those that stick to the facts.

Figure 8 is the average sentiment and subjectivity of the tweets made by the top 20 followed users. In the context of tweets, sentiment is defined as an attitude, thought, or judgment. In Textblob, the score ranges from 0 for very objective and 1 for heavily subjective or opinionated.

Time Series Point Analysis

We further explored the data and annotated the series of tweets using the hashtags and their popularity. Figure 9 shows the series of tweets and the popularity of the various hashtags. Using the power of hindsight and the Wikipedia of US current events, we can see what some notable spikes are related to:

- August 4, 2016: spike of hash tag #obamaswishlist which were posts about fanciful and perceived hypocritical items Obama “wanted”
- August 17, 2016: spike of the hash tag #trumpsfavoriteheadline which are Tweets about sardonic headlines that Donald Trump would endorse
- September 28, 2016: #ihavearighttoknow movement by fake accounts to know what Hillary Clintons emails were
- October 5, 2016: #ruinadinnerinonephrase was actually seen as both politically-backed and non-politically-backed with some referencing it to Hillary Clinton while others made memes out of the hashtag
- October 17, 2016: #makemehateyouinonephrase, another hash tag movement that was seen as either part of a meme culture or part of the political systems

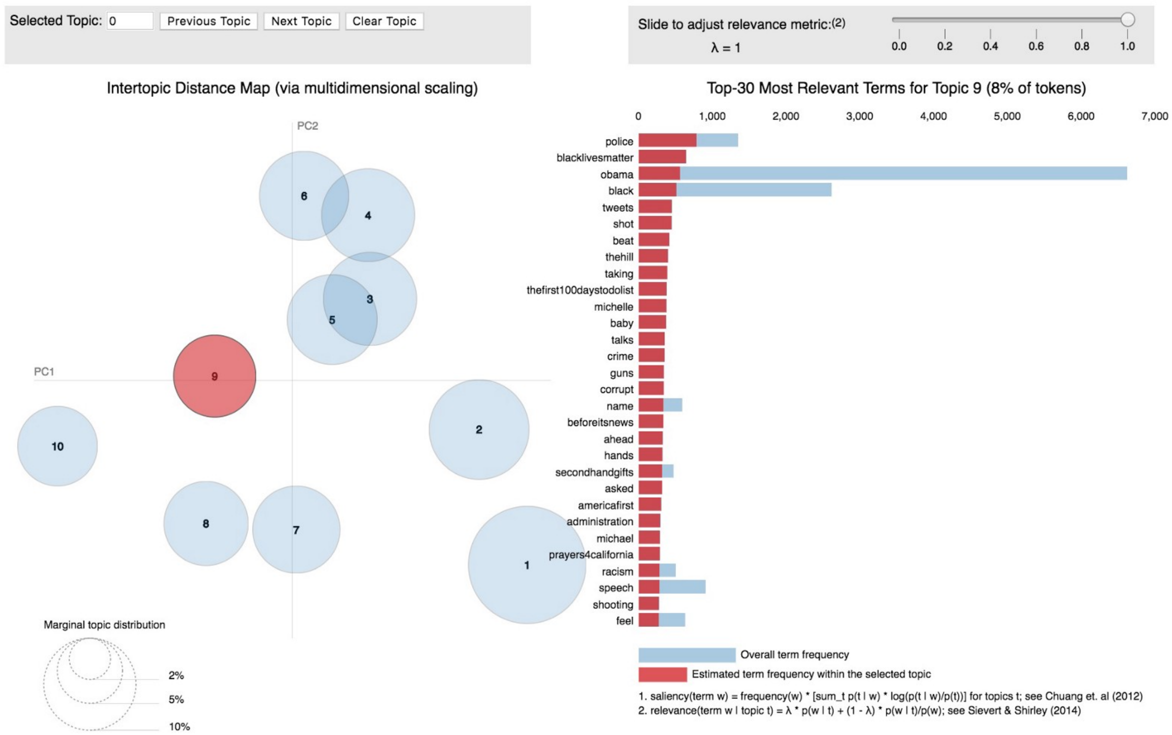

Figure 6. *Topics in the tweets of the fake users. The left half shows the various topics and the right half of the figure shows a histogram of the terms in a user selected topic. We picked Topic 9 on the left and related terms for that topic such as Black Lives Matter, Obama, AmericaFirst, and so on are shown on the right. We used the LDAVis [20] tool to analyze the bios.*

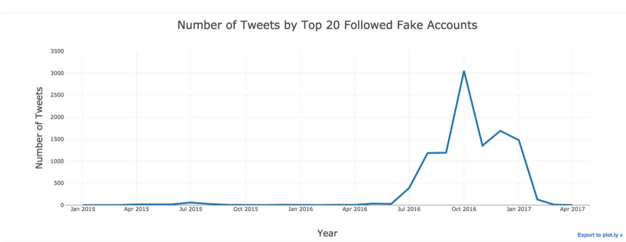

Figure 7. *Number of Tweets by the top 20 Fake Accounts. We can see that the volume of the tweets was highest during the run up to the Elections which was followed by a slow demise to an almost inactive state in 2017.*

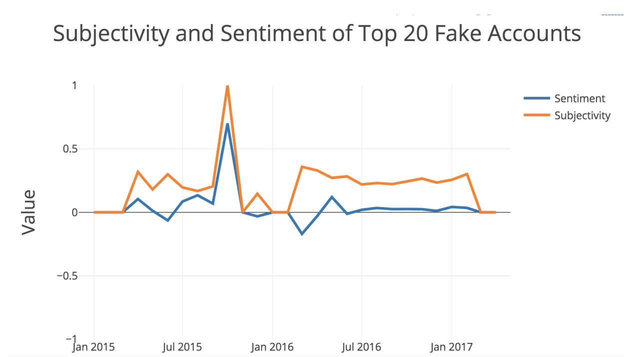

Figure 8. *Average Sentiment and Subjectivity of Top 20 Followed Users. The positive sentiment and subjective tweets were seen during crucial times in the election cycle. The first big spike was seen during the* ***first*** *Republican debate on August 6, 2015. The second big spike is just before and on March 1st, 2016 which was Super Tuesday when 11 states voted in their primaries for their parties candidates.*

- November 14, 2016: #reallifemagicspells used in reference with black lives matters and Trumps family
- December 7, 2016: #idrunforpresidentif "Id known I needed literally zero experience" and other sardonic comments about the presidential election

Perhaps coincidentally or not, the initial spikes were all related to fake accounts simultaneously using hashtags to mock presidents and presidential candidates. The tweets were clearly politically-based with the name drops to actual candidates. However, as time progressed, the distinguishing factor between these tweets became less obvious, as the fake accounts used actual popular hashtags that were not clearly political. Additionally, the tweets seemed to be initially aimed at all the candidates rather than one particular candidate until Trump was actually elected, at which point these fake accounts joined the popular hashtags attacking Trump.

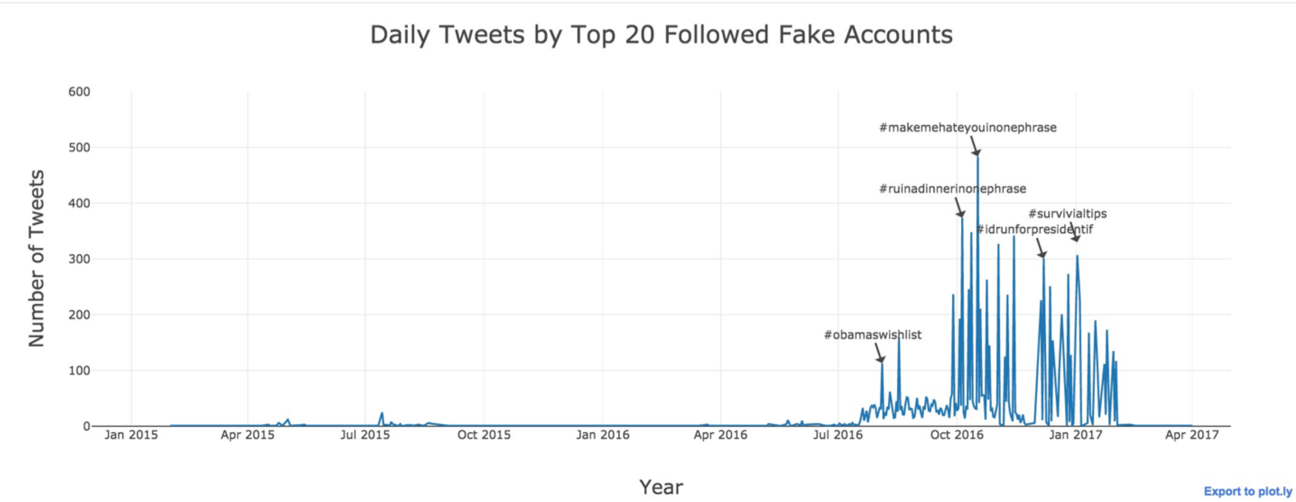

Figure 9. *Time Series Tweets with Hash Tags. This figure shows when the top 20 most influential accounts tweeted and has been annotated with popular hashtags coinciding with the spikes. The fake accounts joined in with the popular hashtags and injected propaganda in them.*

Conclusion

From our analysis, we learned that the fake accounts disguised themselves as (1) average Americans, (2) news sites with metropolitan names, or (3) international names that describe themselves with relatable topics such as political and religious beliefs. We also found that they achieved their objective of influencing Twitter users by posting polarizing tweets at opportunistic times such as the weekends when scandals and large announcements occurred. Finally, they grew sentient of their obvious posts by subtly joining trending hashtags and injecting propaganda within it.

References

[1] Saeed Abdullah, Elizabeth L Murnane, Jean MR Costa, and Tanzeem Choudhury. Collective smile: Measuring societal happiness from geolocated images. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 361–374. ACM, 2015.

[2] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[3] Julio Borges, Henrik Hain, Simon Sudrich, and Michael Beigl. Event detection for smarter cities. In *2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. IEEE, 2017.

[4] Donn Byrne. An overview (and underview) of research and theory within the attraction paradigm. *Journal of Social and Personal Relationships*, 14(3):417–431, 1997.

[5] Siming Chen, Lijing Lin, and Xiaoru Yuan. Social media visual analytics. In *Computer Graphics Forum*, volume 36, pages 563–587. Wiley Online Library, 2017.

[6] Siming Chen, Xiaoru Yuan, Zhenhuang Wang, Cong Guo, Jie Liang, Zuchao Wang, Xiaolong Luke Zhang, and Jiawan Zhang. Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data. *IEEE transactions on visualization and computer graphics*, 22(1):270–279, 2016.

[7] Nicholas Diakopoulos, Mor Naaman, and Funda Kivran-Swaine. Diamonds in the rough: Social media visual analytics for journalistic inquiry. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, pages 115–122. IEEE, 2010.

[8] Sri Krisna Endarnoto, Sonny Pradipta, Anto Satriyo Nugroho, and James Purnama. Traffic condition information extraction & visualization from social media twitter for android mobile application. In *Electrical Engineering and Informatics (ICEEI), 2011 International Conference on*, pages 1–4. IEEE, 2011.

[9] Andrea Ferracani, Daniele Pezzatini, Lea Landucci, Giuseppe Becchi, and Alberto Del Bimbo. Separating the wheat from the chaff: Events detection in twitter data. In *Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing*, page 14. ACM, 2017.

[10] Prasanna Giridhar, Shiguang Wang, Tarek Abdelzaher, Tanvir Al Amin, and Lance Kaplan. Social fusion: Integrating twitter and instagram for event monitoring. In *Auto-*

nomic Computing (ICAC), 2017 IEEE International Conference on, pages 1–10. IEEE, 2017.

[11] Guardian. Twitter admits far more russian bots posted on election than it had disclosed. `https://www.theguardian.com/technology/2018/jan/19/twitter-admits-far-more-russian-bots-posted-on-election-than-it-had-disclosed`, 2018.

[12] Nadav Hochman and Lev Manovich. Zooming into an instagram city: Reading the local through social media. *First Monday*, 18(7), 2013.

[13] Steven Loria. Textblob - simplified text processing. `https://textblob.readthedocs.io/en/dev/`, 2017.

[14] Deepa Mallela, Dirk Ahlers, and Maria Soledad Pera. Mining twitter features for event summarization and rating. In *Proceedings of the International Conference on Web Intelligence*, pages 615–622. ACM, 2017.

[15] Jose Maria Balmaceda, Silvia Schiaffino, and Daniela Godoy. How do personality traits affect communication among users in online social networks? *Online Information Review*, 38(1):136–153, 2014.

[16] Fabio Miranda, Harish Doraiswamy, Marcos Lage, Kai Zhao, Bruno Gonçalves, Luc Wilson, Mondrian Hsieh, and Cláudio T Silva. Urban pulse: Capturing the rhythm of cities. *IEEE transactions on visualization and computer graphics*, 23(1):791–800, 2017.

[17] NBC News. Twitter deleted 200,000 russian troll tweets. read them here. `https://www.nbcnews.com/tech/social-media/now-available-more-200-000-deleted-russian-troll-tweets-n844731`, 2018.

[18] Team NLTK. NLTK Naïve Bayes Classifier. `https://www.nltk.org/`, 2017.

[19] Joanna Schug, Masaki Yuki, Hiroki Horikawa, and Kosuke Takemura. Similarity attraction and actually selecting similar others: How cross-societal differences in relational mobility affect interpersonal similarity in japan and the usa. *Asian Journal of Social Psychology*, 12(2):95–103, 2009.

[20] Carson Sievert and Kenneth Shirley. Ldavis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces*, pages 63–70, 2014.

[21] New York Times. Facebook and cambridge analytica: What you need to know as fallout widens. `https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html`, 2018.

[22] Chaolun Xia, Jun Hu, Yan Zhu, and Mor Naaman. What is new in our city? a framework for event extraction using social media posts. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 16–32. Springer, 2015.

[23] Chaolun Xia, Raz Schwartz, Ke Xie, Adam Krebs, Andrew Langdon, Jeremy Ting, and Mor Naaman. Citybeat: real-time social media visualization of hyper-local city data. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 167–170. ACM, 2014.

[24] Panpan Xu, Yingcai Wu, Enxun Wei, Tai-Quan Peng, Shixia Liu, Jonathan JH Zhu, and Huamin Qu. Visual analysis of topic competition on social media. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2012–2021, 2013.