Rejecting the Null Hypothesis of Apathetic Retweeting of US Politicians and SPLC-defined Hate Groups in the 2016 US Presidential Election

Raazesh Sainudiin
Department of Mathematics
Uppsala University
Uppsala, Sweden
raazesh.sainudiin@math.uu.se

Kumar Yogeeswaran
Department of Psychology
University of Canterbury
Christchurch, New Zealand
kumar.yogeeswaran@canterbury.ac.nz

Kyle Nash
Department of Psychology
University of Alberta
Edmonton, Canada
knash@ualberta.ca

Rania Sahioun
Department of Psychology
University of Canterbury
Christchurch, New Zealand
rouna79@yahoo.com

***Abstract*—We characterize the Twitter networks of both major presidential candidates, Donald Trump and Hillary Clinton, with various American hate groups defined by the US Southern Poverty Law Center (SPLC). We further examined the Twitter networks for Bernie Sanders, Ted Cruz, and Paul Ryan, for 9 weeks around the 2016 election (4 weeks prior to the election and 4 weeks post-election). By carefully accounting for the observed heterogeneity in the Twitter activity levels across individuals under the null hypothesis of apathetic retweeting that is formalized as a random network model based on the directed, multi-edged, self-looped, configuration model, our data revealed via a generalized Fisher's exact test that there were significantly many Twitter accounts linked to SPLC-defined hate groups belonging to seven ideologies (Anti-Government, Anti-Immigrant, Anti-LGBT, Anti-Muslim, Alt-Right, Neo-Nazi, and White-Nationalist) and also to @realDonaldTrump relative to the accounts of the other four politicians. The exact hypothesis test uses Apache Spark's distributed sort and join algorithms to produce independent samples in a fully scalable way from the null model.**

***Index Terms*—Donald Trump, Twitter, 2016 US Presidential election, US hate groups, configuration model, scalable generalized Fisher's exact test, Apache Spark**

I. Introduction

During the 2016 US Presidential election, considerable debate emerged about whether Donald Trump has unique appeal to people with hateful and prejudiced views. Some argued that Trump appealed to Whites threatened by increased cultural diversity and anti-minority sentiments [1], while others focused more on his appeal to the politics of hope and identity [2]. Did Trump's candidacy possess unique appeal for people espousing hateful ideologies? Or did he have the same appeal as other prominent politician during the election? We address this question using nearly 22 million communication events on Twitter that characterized the social media networks of Donald Trump, Hillary Clinton, Bernie Sanders, Ted Cruz, Paul Ryan, and several American hate groups, as defined by the US Southern Poverty Law Center (SPLC) [3]; hereafter referred as hate groups.

Using an exact test, we address this question by attempting to reject the null hypotheses of *apathetic retweeting* where each user is apathetic (ie., indifferent with a lack of interest or concern[1]) about who they retweet while preserving the number of times they retweet others and the number of times others retweet them. We formalize this null model as the *apathetic retweet network* from the configuration network model [4] (see [5] for an overview of models for stochastic social dynamics).

II. Data, Design, Models and Methods

A. Data and Experimental Design

1) Communications in Twitter: Twitter, a popular micro-blogging service, provides an observable social network with millions of users [6]. Twitter allows users to communicate or update their status in many ways. One can post a message called a *tweet* that is no more than 140 characters in length (in 2016), *follow* another Twitter user, and receive the status updates of every user they follow. A tweet that is shared publicly with one's own followers is known as a *retweet*.

2) Dataset: Our dataset comprises of 21,749,868 communication events in Twitter over a total of 9 weeks centered around the US 2016 election. The tweet IDs are available from https://tinyurl.com/y8lamxzx. The dataset was collected in our designed experiment using Twitter's streaming and REST APIs [7] that were extended for Apache Spark [8], a unified engine for big data processing. Each of these communication events were parsed and classified into seven types of events using SparkSQL modules.

Streaming data collector: Over 17 million events were collected from Twitter's public streams by directly tracking communications related to the Twitter accounts of the five political candidates and 52 hate group accounts. The SPLC database does not necessarily consider all groups or individuals on its list as violent or engaged in criminal activities, but

R.Sai. was partly supported by the chaire Modélisation Mathématique et Biodiversité of Veolia Environnement–École Polytechnique–Museum National d'Histoire Naturelle–Fondation X. Distributed cloud computing of Twitter data by R.Sai. was supported by Databricks Academic Partners Program.

[1] "Apathy." Merriam-Webster.com. Accessed July 10, 2018. https://www.merriam-webster.com/dictionary/apathy.

rather identifies any group or individual "whose beliefs or practices attack or malign an entire class of people, typically for their immutable characteristics" [3]. The database does not include foreign hate groups or extremist groups such as ISIS, Al Qaeda, or Boko Haram, as its focus is on American hate groups. The SPLC database categorizes the activities of individuals and groups into several ideologies including Neo-Nazi (e.g., National Socialist Movement), Black-Separatist (e.g., Nation of Islam), Anti-LGBT (e.g., Westboro Baptist Church), Anti-Immigrant (e.g., American Border Patrol), Anti-Muslim (e.g., ACT for America), Alt-Right (e.g., American Renaissance), Anti-Government (e.g., Oath Keepers), and White-Nationalist (e.g., Aryan Brotherhood). Because only 78% of the SPLC identified hate groups had a valid Twitter account that we could track, our study is limited to a further subset of 52 user accounts who were active in the Twitter public streams. Thus, our approach is not exhaustive, in terms of being able to track every account of each hateful ideology, but is nonetheless representative of the public Twitter activity[2].

Retrospective data augmentation: The remaining 5 million events were obtained retrospectively using the REST API as follows. Due to most retweets being an immediate reaction to a tweet that one finds interesting or concurs with [10], nearly 7 million of the 10.5 million retweets in the 17 million events collected from the public streams happened within the same day of the original tweet and over 98% of the retweets happened within a week of the original tweet. Furthermore, since our tweet collector is only recording events that are directly related to one of our tracked users, we do not know of any other Twitter interactions by those who retweeted one of our tracked users. Thus, to better understand the recent retweet behavior of at least some of the politically active Twitter users with other non-tracked users on Twitter, we focus on October 19 2016, the day of the 3rd US Presidential debate, and obtain a seed set of users who retweeted either @HillaryClinton or @realDonaldTrump on this day. The communication intensity reached over 120 events per second around the debate in our streaming data collector. Our seed set is made up of a random sample of about a third of all users (including all verified and geo-enabled accounts) who retweeted either Clinton or Trump on October 19 2016 and thus constitute an evenly represented sample of politically active Twitter users from the two parties. For each user in the seed set, at the end of the 9 week period, we added all the retweets from their 200 most recent status updates that occurred in the 9 week period. This strategy involves a breadth-first expansion about the seed set of users in the much larger retweet network on Twitter as it allows us to expand our 9-week-long retweet network by focusing on the recent retweet timelines of those who retweeted either of the two final presidential candidates during the day of the last debate. Crucially, this augmented data added another 0.3 million users to our network, increased the number of retweet events from 10.5 million to 13.7 million and made the retweet network into a single connected component.

[2] Although the SPLC also identifies ideologies of the Ku Klux Klan (KKK), Neo-confederates, Racist Skinheads, and Holocaust Deniers, within their database, there was not enough activity on the Twitter accounts of these groups that we were able to track given just a few thousand followers. It may be that these groups have a smaller online presence, or they simply use a different social media platform (including closed discussion forums; see [9]) making it harder to follow as we focused our analyses only on hate groups with public streams.

B. Modeling Retweets

1) Retweet Network: Retweets are a simple and natural signal of *directional concurrence*, i.e. concurrence *with* the user who posted the (original) tweet *by* the user who retweeted it (i.e., between the tweeter and the retweeter), as they express interest in the message and also endorsement and trust in the communicator [6], [11], especially when retweeted multiple times. Retweet analysis via random network models as done in this study circumvents the ambiguity and uncertainty associated with statistical algorithms in natural language understanding (NLU) that will be required when working with (a) *quoted tweets* where one is allowed to add a comment to the retweet, (b) *reply tweets* or (c) *reply of quoted tweets* where one can reply in possible disagreement.

The number of retweets per day captured from the public stream during the 9 week period reached over 1 million during the day of the third US Presidential debate and the days leading up to the election. The proportion of retweets is known to vary greatly depending on the features of the Twitter subnetwork under study [10]. There were over 13.7 million retweets (63.1%) of an original tweet and over 2.7 million (12.6%) original tweets in our dataset. We ignore the remaining 25% of the events that require further NLU and focus our analysis instead on the retweet network obtained from 75% of all events in our dataset as explained in the next section.

We allowed the set of users who tweet and retweet, i.e., tweeters and retweeters, to form the nodes of the retweet network. Each retweet was allowed to represent a directed edge or arc from its tweeter to its retweeter, i.e. the tweeter-retweeter pair, as a signal of directional concurrence in the retweet network. Because one can retweet more than one tweet posted by any user including oneself, we allowed for parallel or multiple edges between the same pair of users including self-loops, i.e. edges from and to the same user. There were over 4.4 million unique tweeter-retweeter pairs out of 2.5 million unique users representing over 16.4 million tweets and retweets in our dataset with over 21.7 million communication events during the 9-week period around the US presidential election. Mathematically, the retweets in our dataset are represented by a *directed multi-edged self-looped network*. Thus the *out-degree* (number of outgoing edges) and *in-degree* (number of incoming edges) for a user in the retweet network gives the number of times that the user is retweeted by others and the number of times the user retweets others (including oneself), respectively. Similarly, the number of distinct users who are retweeted by the user is given by their *in-nbhd* and the number of distinct retweeters of the user is given by their *out-nbhd*.

The observed retweet network is highly heterogeneous and largely dominated by the two presidential candidates as depicted in Table I. Due to our Twitter collector's design with retrospective data-augmentation, our network forms a single connected component when viewed with undirected edges.

TABLE I
RETWEET NETWORK STATISTICS OF THE FIVE POLITICAL ACCOUNTS

Politician	in-degree	in-nbhd	out-degree	out-nbhd
Donald Trump	40	12	5,952,257	958,262
Hillary Clinton	225	121	2,774,111	943,995
Bernie Sanders	107	62	762,209	356,718
Paul Ryan	769	158	68,973	28,902
Ted Cruz	322	189	49,479	27,663

2) Apathetic Retweet Network Model: The *configuration model* for directed networks [4] is a random network model that produces samples uniformly from the set of networks that preserve the in-degree and out-degree of each node in given observed network.

By representing each retweet as a directed edge between the author of the tweet and its retweeter, we define the null apathetic model for a random retweet network as the *directed, multi-edged, self-looped configuration model*, whereby each user apathetically or non-preferentially retweets another user, by choosing uniformly at random from the set of all users, while preserving their observed in-degree and out-degree. However, the model frees up information on the users involved in retweeting, i.e. who retweets whom, provided the in-degree and out-degree of each user is identical to that of the observed network. This null model thus preserves the innate tweet and retweet rates of each user while modeling complete apathy or indifference for who one chooses to retweet.

3) Cut, Permute and Rewire: CUTPERMUTEANDREWIRE is a Monte Carlo algorithm to generate independent sample networks from the directed multi-edged self-looped random configuration model, i.e. our null random network model for apathetic retweeting. CUTPERMUTEANDREWIRE is a distributed, scalable, and fault-tolerant version of the standard construction involving random pairings of out-bound and in-bound half-edges [4] through the following three steps: (i) *cutting* the directed edges representing the retweets in our observed retweet network into out-bound and in-bound half edges, (ii) *permuting* the in-bound half edges by sorting them according to pseudo-random numbers that are generated and associated with them and (iii) *rewiring* the original out-bound half edges with the permuted in-bound half edges using a distributed join. Note that the in-degree and out-degree of each node in the observed retweet network is preserved after the three steps by construction.

By taking advantage of the fastest available distributed sorting and optimized distributed joining algorithms in Apache Spark [8], CUTPERMUTEANDREWIRE can produce independent sample networks with tens of millions of retweets or edges in a small Apache Spark cluster over six commodity compute nodes. The null distribution of any test statistic of the retweet network can be directly obtained from applying it to each independent Monte Carlo sample from our scalable fault-tolerant randomized algorithm with probability $1/(\text{number of retweets})!$ under the null model. By comparing the observed test statistic to Monte Carlo samples from the null distribution, one can directly obtain consistent estimates of the p-value in order to attempt to reject the null hypothesis of apathetic retweeting in favor of the alternative hypothesis of non-apathetic retweeting in the framework of generalized Fisher's exact test.

RESULTS & DISCUSSION

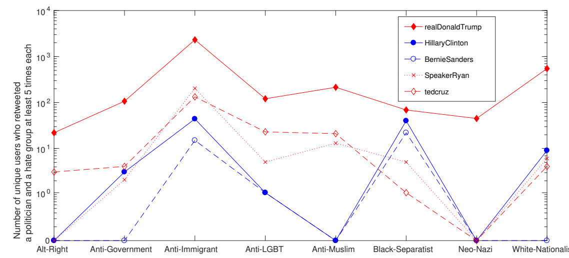

Fig. 1. Number of unique users who retweeted a politician and a hate group at least five times each (Note: The y-axis is in log-scale in powers of 10).

Given the observed heterogeneity in the activity levels of Twitter users, including the most influential user @realDonaldTrump, who was more active than the other four politicians with over twice as many retweets by about the same number of retweeters as the next most influential user @HillaryClinton (Table I), any hypothesis test needs to control for this heterogeneity. Our novel approach uses retweet networks sampled independently from the null apathetic retweet network model in order to obtain the null distribution of any test statistic while preserving the observed differences in the activity levels, in terms of the observed in-degree and out-degree of each user in the retweet network. Note that we do not have multiple testing issues when attempting to reject the null hypothesis of apathetic retweeting in favor of its complementary alternative hypothesis of non-apathetic retweeting since we obtain independent Monte Carlo samples from the joint distribution of the statistics under the null model to estimate the acceptance and rejection regions at the significance level of 0.001.

For a robust test statistic we looked one degree deeper into the retweet network and obtained the number of Twitter users who retweeted both a politician and a specific hate group at least five times. This provided a more conservative estimate of the number of users retweeting hate groups or politicians as it would only include those actively retweeting both. The observed frequency of such users is shown in Fig. 1 and Table II along with their 0.1% marginal intervals of the acceptance region for the null hypothesis of apathetic retweeting. Since the observed frequencies jointly lie well outside the acceptance region obtained from 1000 Monte Carlo samples drawn from the null model using the CUTPERMUTEANDREWIRE algorithm, we reject the null hypothesis of apathetic retweeting

TABLE II
OBSERVED FREQUENCY OF DISTINCT USERS WHO RETWEETED A POLITICIAN AND AN ACCOUNT WITHIN A HATE GROUP AT LEAST 5 TIMES EACH

Ideology	Politician				
	Donald Trump	Hillary Clinton	Bernie Sanders	Paul Ryan	Ted Cruz
	observed test statistic: marginal interval for the region of acceptance at 0.001 significance level				
Anti-Government	*107 : (0, 1)	3 : (0, 3)	0 : (0, 1)	*2 : (0,1)	*4 : (0, 1)
Anti-Immigrant	*2314 : (375, 498)	$^\circ$44 : (373, 492)	$^\circ$15 : (369, 485)	*204 : (47, 95)	*133 : (18, 54)
Anti-LGBT	*121 : (0, 4)	1 : (0, 4)	1 : (0, 4)	*5 : (0, 3)	*23 : (0, 3)
Anti-Muslim	*215 : (0, 3)	0 : (0, 3)	0 : (0, 3)	*13 : (0, 3)	*21 : (0, 3)
Neo-Nazi	*45 : (0, 1)	0 : (0, 1)	0 : (0, 1)	0: (0, 1)	0: (0, 1)
White-Nationalist	*548 : (0, 12)	9 : (0, 10)	0 : (0, 10)	6: (0, 8)	4: (0, 7)
Black-Separatist	$^\circ$69 : (653, 811)	$^\circ$40 : (649, 808)	$^\circ$22 : (645, 801)	$^\circ$5 : (72, 128)	$^\circ$1 : (28, 66)
Alt-Right	*22 : (0, 0)	0 : (0, 0)	0 : (0, 0)	0 : (0, 0)	*3 : (0, 0)

in favor of non-apathetic retweeting with a p-value less than 0.001.

When the observed counts are in the rejection region by falling below or above the respective marginal intervals of the acceptance region, we can obtain statistically meaningful insights of the observations under the favored alternative hypothesis of non-apathetic retweeting. However, such insights need further models and tests for a refined and rigorous understanding within the space of alternative hypotheses.

In Table II the observed frequencies that lie in the rejection region above and below the corresponding marginal intervals for the acceptance region are indicated by * and $^\circ$, respectively. Thus, for example, a statistically significant frequency of users have retweeted Trump, Ryan or Cruz and Anti-Immigrant groups at least five times each in a non-apathetic and preferentially approving manner since these frequencies are above their null intervals, i.e. the marginal intervals of the acceptance region for the null model of apathetic retweeting at a significance level of 0.001. These same frequencies are below the null intervals for Clinton and Sanders in favor of the non-apathetic retweeting alternative hypothesis, but in a preferentially disapproving manner. Only Trump has a significant frequency of retweeters who also frequently retweet Anti-Government, Anti-Immigrant, Anti-LGBT, Anti-Muslim, Neo-Nazi, White Nationalist and Alt-Right hate groups in a non-apathetic and preferentially approving manner. Note that the Black-Separatist hate group forms a "natural control" as their retweeters non-apathetically retweet all five politicians in a preferentially disapproving manner.

Our results are robust to changes in the threshold number of retweets being at least 5, albeit at the expense of fewer observations for larger thresholds. This is because the pattern of user frequencies across hateful ideologies and the five politicians shown in Fig. 1 is preserved across other thresholds. The results are also robust when the test statistic is an expectation of the frequencies taken over all threshold levels greater than 1. We chose the naturally interpretable threshold of at least 5 retweets over the 9-week period of observation in Fig. 1 and Table II to select users with more than one retweet per fortnight, on average.

Thus, we characterize the Twitter networks of five major politicians and various American hate groups (as defined by the US Southern Poverty Law Center) for 9 weeks around the 2016 election. By carefully accounting for the observed heterogeneity in the Twitter activity levels across individuals under the null hypothesis of apathetic retweeting that is formalized as a configuration model, our data revealed via a scalable generalized Fisher's exact test that there were significantly many Twitter accounts linked to US hate groups belonging to seven ideologies (Anti-Government, Anti-Immigrant, Anti-LGBT, Anti-Muslim, Alt-Right, Neo-Nazi, and White-Nationalist) and also to @realDonaldTrump relative to the accounts of the other four politicians. Thus, our approach towards scientific understanding of social media interactions from statistical experiments with big data uses distributed and fault-tolerant algorithms through Apache Spark [8] in order to perform exact statistical hypothesis tests with an interpretable and scalably samplable null random network model.

REFERENCES

[1] B. Major, A. Blodorn, and G. Major Blascovich, "The threat of increasing diversity: Why many White Americans support Trump in the 2016 presidential election," *Group Proc Intergr Rel*, 2017, in press.

[2] S. D. Reicher and S. A. Haslam, "Trumps appeal: What psychology tells us," *Scientific American Mind*, March 2017.

[3] Southern Poverty Law Center, "Hate map," *SPLC*, 2016, october 11, 2013 4:00 AM, Available from https://www.splcenter.org/hate-map. Accessed on May 28, 2017.

[4] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Random graphs with arbitrary degree distributions and their applications," *Phys. Rev. E*, vol. 64, p. 026118, Jul 2001.

[5] D. Aldous, "Interacting particle systems as stochastic social dynamics," *Bernoulli*, vol. 19, no. 4, pp. 1122–1149, 09 2013.

[6] B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury, "Twitter power: Tweets as electronic word of mouth," *J Amer Soc Info Science Tech*, vol. 60, no. 11, pp. 2169–2188, 2009.

[7] "Twitter developer documentation," https://dev.twitter.com/docs, 2017, accessed on May 17 2017.

[8] M. Zaharia, R. S. Xin, P. Wendell, T. Das, M. Armbrust, A. Dave, X. Meng, J. Rosen, S. Venkataraman, M. J. Franklin, A. Ghodsi, J. Gonzalez, S. Shenker, and I. Stoica, "Apache spark: A unified engine for big data processing," *Commun. ACM*, vol. 59, no. 11, pp. 56–65, oct 2016.

[9] T. Martin, "Dissecting Trump's most rabid online following," *FiveThirtyEight*, 2017, available from https://fivethirtyeight.com/features/dissecting-trumps-most-rabid-online-following/. Accessed on June 14, 2017.

[10] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" in *Proceedings of the 19th International Conference on World Wide Web*, ser. WWW 10. New York, NY, USA: ACM, 2010, pp. 591–600.

[11] P. Metaxas, E. Mustafaraj, K. Wong, L. Zeng, M. O'Keefe, and S. Finn, "What do retweets indicate? results from user survey and meta-review of research," in *International AAAI Conference on Web and Social Media*. ACM, 2015.