“The devil is in the details : Studying the Influence and Content Diffusion Dynamics of Social Bots During 2014 Italian Mayoral Elections”

Pujan Paudel
The University of Southern Mississippi, USA

Abstract
Twitter bots have been exercised by political parties and state-level agencies in many contexts. From state-sponsored trolls to political amplifiers, bots have made their mark in the elections. But, the question of how effective and successful strategies of these bots could be in comparison to humans in the same election is an open question. In this work, we study the Twitter bots in the Mayoral Election of Italy in 2013 and compare them with human users activity. Our analysis of the echo-chamber effect, the emotional cost of information diffusion, and mechanism of information spreading reveals us the bots being able to reach wider ideology of political ideology by tweeting with lesser emotional cost in their tweets. We also conduct meme analysis and longitudinal analysis of user roles where the Twitter bots have an edge on showing more effective tweeting and diffusing behaviours. We then discuss the possible implications of this effective behaviour of bots on multiple dimensions and how twitter bots could be deployed in jeopardizing the political process.

Introduction
Political bots in Twitter have evolved their strategies and forms of attack with time, and they can no more be marked off as merely content spammers or conversation polluters. The botmasters have adapted their sock puppets to be capable of suppressing flow of democratic voices through collective demobilization of social botnets and change perspectives of public on social discussions, actors and even the entire political regime through the volumetric mirage of automated influence.

Literature Review

1. Twitter bots in political contexts
Automated agents, or bots in Twitter have been identified as key players acting on the platform, taking part in various social movements across different countries and political contexts. Evidence of bots and their high usage activity have been highlighted in works of Bessie & Ferrara [1] in the US elections and Howard and Kollanyi [2] in the Brexit debate. Previous works have identified varied usage of bots in Twitter as sockpuppets in Syrian civil war [5], political propaganda spreaders in Ukraine Russian conflict [7] and petition hijackers in the Brexit referendum[9]. Recent studies have also highlighted alarming behaviors of social bots in being the key drivers of Macronleaks

campaign [10] and effective spreaders of low credibility content [13]. Automated accounts on Twitter have been deployed from the state or government level itself in multiple instances. Militaries and state sponsored firms were found to be spreading political propaganda in Venezuela [3] and massive troll armies backed by Russian state were allegedly used to promote pro -state ideals [4]. The misuse of bots have also been observed in Asian political contexts, as the Chinese political bots were found to be active against anti-government public campaigns in Tibet and Taiwan [4], and it was discovered the South Korean secret service to be using cyborg trolls in astroturfing behaviors at the 2012 elections. Multiple sources have also revealed the involvement of Mexican government behind deployment of Twitter botnet to suppress public discussion and distort communication of anti government discussions [4].

2. Strategies of political bots usage

The strategies of the automated agents discussed above vary on different levels, based on their primary objective of deployment. Relatively simple applications of Twitter bots have been on the line of using them to statistically pad the social network size of political actors [4]. The prominent usage of political bots have been motivated towards astro-turfing behaviors . From creating false effects of group consensus about particular political candidates ,and political agendas [12] to systematically producing more positive content in support of a political candidate [1], the bots were employed to distort the volumetric as well as emotional social presence of political actors and parties. The increased social sensing power of the automated agents, through advancement in Big Data methods have given the bots the ability to model their profile characteristics and the information they produce to be highly deceptive and human like. Those bots have been successful in targeting reply networks of genuine human users within the same political spectrum of their political campaigns and generate long retweet cascades [6]. The findings of [2] and [1] have revealed bots to be taking part in both sides of the political discussion. In multiple case studies across different political contexts, [4] and [5] have pointed out the coordinated usage of the Twitter botnet to flood Twitter stream with pro-campaign hashtags, and spoil meaningful,conversational hashtag with spam content to be a widely used strategy.

3. Influence of political Twitter bots

The political bots in Twitter have demonstrated high effectiveness in their strategies at multiple instances, demonstrating their ability to influence social processes extending from cyber space to the socio-physical levels. The usage of social bot armies to retweet pro-candidate and pro-party content in South Korean elections of 2012 was found to have heavy influence in swaying public opinion in favour of Park Geun-hye, the eventual winner of the elections [2] . It can be argued that that the bots played an influential role in the political discussion surfacing the US presidential election and

Brexit debate, as the results of the events were different from their initial polls and forecasts and the discussions involved high volume of activity on Twitter with involvement from automated agents. [1] and [2]. The bots used in US Presidential election were able to occupy influential positions in communication networks, and gather equal shares of retweet as genuine human users [1]. In terms of deception , the social bots have been proved to be highly effective as the study of [16] demonstrated that the users were unable to differentiate between the tweets coming from bots or humans signifying the huge potential of political bots to have detrimental effects on the ethical and democratic aspects of elections and political processes. The bots were highly successful in affecting the dynamics of the discussion networks as they were able to influence measurements of influence like Klout, Kred, and Retweet Rank [17] as well as dominating the network in terms of engagement and betweenness centrality[14]. Even the most simplest of acts of retweeting political hashtags by the political bots resulted in amplification of political voices of less prominent actors [14]. The potential for extreme influence of coordinated social bots acting together was highly emphasized by the work of [6] as they demonstrated a very small botnet of 9 actors, set up in 13 minutes being able to generate retweet cascades from genuine users. In perspectives of fairness of information, the bots were found to be highly influential as the collective action of actors in a botnet could raise retweeted URLs and slogan hashtags to the top of Google Search results [12].

2. Datasets

For our study of political social bots, we took the dataset of bots from the work of [20]. The dataset consists of a novel group of social bots which were deployed by one of the runner-up parties of the 2014 Mayoral election of Rome employed through a social media marketing firm. The manually annotated dataset of bots displayed very human like characteristics in profile and content based attribute. In our recent work [19], we discovered how those bots have evolved over the traditional political bots network wise, communication wise and behavior wise. We were further motivated by the complex and human like resemblance shown by the political social bots to compare their activities with human users taking part in the same election. Since the bots in the datasets interchangeably mixed their tweets with content the current public was discussing, we removed such noise from the dataset of [20] by removing tweets not containing any of the politically relevant keywords. We identified the politically relevant keywords the bots were engaged in by initially creating a seed set of hashtags used in the re-tweets of the political accounts whom the bots we promoting. We then expanded our keywords by identifying the hashtags that co-occured with the seed hashtags more than 10 times. We translated the hashtag to make sure that our keyword based approach for removing the noise is indeed consistent. The top keywords which we used to identify the politically relevant conversation is listed in *Table I*. To identify human users to compare the activity of bots, we needed to identify Twitter users who took part in the political discussion during the surfacing event of 2014 Mayoral elections. Based on the work of [21], the conversations on twitter can be collected using 3 different strategies , i) using hashtags ii) using random stream iii) using official set of dictionary users known to be talking on the topic. We applied the methods i) and ii) to collect users and tweets for constructing the human user dataset as we were not able to use the random stream of Twitter to get archival data back from 2014. For the "hashtag" based collection approach ,we collect all the tweets containing either of the keywords in *Table I* during the identical timeline of the tweets filtered from the bots during 2013-09-15 to 2014-11-30. It can be observed in *Table I* that we cover a wide variety of hashtags related to the primary political party leaders under concern, the opposition party, and external affairs to encompass the different dimensions of election conversations taking part on the social network. We used the GetOldTweets-Python [22] tool to collect the data. We then removed the tweets and users who occurred less than 10 times in our collected sets to remove possible hashtag hijackers and content spammers. We also removed all the users who followed any of the bot accounts from the botnets, to try to capture more genuine users and their activities. We refer to the data collected through this approach as *users-samples-1*. For the official set data collection strategy, we initially identified the top 5 political candidates which were retweeted the most by the twitter bots. We then expanded through all the retweeters of the 5 political accounts, during the identical period and eliminated verified accounts and also the accounts with less than 10 retweets originating from the political accounts. This approach gives us a relevant list of human users who were tweeting on the same side of the political spectrum as the political bots during the election period. We also removed the accounts who followed either of the botnet accounts. We refer to the data collected through this approach as *users-samples-2*. We denote the combined datasets (*users-samples-1* and *users-samples-2*) with *users-samples-combined*. Reconstructing the conversation landscape of an event in the past completely would be a very difficult task, but we believe the two approaches we use provide us a close approximation of the discussions. The tweets collected through the official strategy somewhat mitigate the lack of tweets that could have been collected from random stream, as it has been studied that the "streaming strategy" is most similar to the "official set strategy" in terms of relevance and frequency. Our approach also minimizes the drawback of "official set strategy" to oversample broadcast accounts reported on [21] .

3. Methodology

a) Extent of Eco Chambers

The presence of eco-chambers in online discussion is an evitable scenario produced as a result of selective exposures of the platforms and ideological segregation of the users on it. [insert ref of Tweeting from Left to Right …]. Prior works [insert the three primary papers] have highlighted the strong presence of echo-chambers in retweet based interaction in tweets discussing political contexts [insert all 3 papers , Garimella , Quantifying ..] We compared the quantitative presence of eco-chambers induced by the bots and human users by studying how politically polarized are the users who spread or retweet the content of them. In the case of bots, we used the bots who tweeted more than 10 tweets originally authored by the bots of the social botnets, to make sure the human users we included in the study are not random content distributors, but occasionally share the posts of the Twitter bots. In the case of human users, we used the *users-samples-2* dataset we collected, to make sure the human users we studied are actively engaged in discussion about the political activity occurring at the same time the botnet was deployed , possibly being with similar political spectrum as of the Twitter bots. We selected the 500 most frequently used URLs from each of the datasets, and filtered 200 of them being news outlets or media portals. We initially used the lists of partisan media outlets compiled by third party organization Media Bias/ Fact Check[28]. Since most of the sources of media outlets were in Italian language and the outlets not being available on the partisan list, we manually annotated the news outlets by verifying their political alignment cross checking with Wikipedia. We labeled the polarity score of news sources falling under left and left center with -1 , right and right center with 1 and the partisan media outlets with a score of 0. An overview of the media outlets and their polarity score is listed in *Table II* . We then assign two different polarity scores to the users in our study, i) Production Polarity : the mean polarity score of the tweets produced by a user ,which contain links to the media outlet of known political polarization. ii) Consumption Polarity : the mean polarity score of the tweets produced by the users who share the contents of the user.

Politicians	Political Party/ Slogans	Opposition Party	External Affairs
#Renzi	#PD	#Brunetta	#Palestina
#cuperlo	#BelloeDemocratico	#Grillo	#Hamas
#gotti	#congressopd	#Lupi	#Gaza

#bersani	#M5s		#IsraelUnderAttack
#letta			#TelAviv

Left Leaning (Polarity : -1)	Right Leaning (Polarity : +1)
ilmanifesto	iltempo
messianicradio	lefigaro
carmillaonline	breitbart
partitodemocratico	ecodibergamo
rifarelitalia	sicilia5stelle
daysofpalestine	conservativescores
lapresse	algemeiner
primariepd2013	ohalright
volkskrant	ilfoglio

Table II : Top Media Outlets and their polarity score assigned

b) Emotional Cost of Information Diffusion

We studied how the degree of influence gained by a tweet varies with the emotional content present on it. We used the tweets by the bots , and the users from our *users-samples-combined* dataset with number of retweets obtained between 5 and 90. For getting numerical score of the multiple dimensions of emotional attributes conveyed by the tweets, we used the Perspective API [24]. We translated the tweets in our datasets using Google Translate API as the tweets were in Italian Language. We used 5 different dimensions of emotional analysis offered by Perspective API , out of the 16 total. The dimensions of emotion we study, alongside the description of the model annotating the score are :

i) Severe Toxicity : rude , disrespectful or unreasonable comments.
ii) Threat : intention to inflict pain, injury , or violence against an individual or group.
iii) Profanity : swear words ,curse words , or other obscene or profane language.
iv) Insult : Insulting , Inflammatory , or negative comment towards a person or a group of people.

v) Identity Attack : negative or hateful comments targeting someone because of their identity.

We analyzed the tweets of the users in bots dataset under further detail by dividing the tweets of the bots into the active tweets , which are still present on the Twitter platform remaining undeleted , whereas deleted tweets , which have been taken down by the platform , alongside the account tweeting it.

c) **Mechanism of Information Spreading**

We studied whether the broadcast model or the viral model dominated the message spreading of the bots and humans in the social space. The viral diffusion model represents interpersonal communication through a chain of individual-to-individual spreading process. Whereas, the broadcast diffusion model closely relates with diffusion patterns of mass media or marketing efforts where there are limited information source disseminating information to a large number of individuals.

We used the tweets by the bots , and the users from our *users-samples-combined* dataset with number of retweets greater than 10, making sure that the tweets were original tweets. We then created an information cascade for all the tweets in our analysis, by using the metadata attached with the tweets , the followers of the users in our study , and the retweet information. The information cascade we constructed represents the path of diffusion of the tweets from the seed node to the message spreaders. To solve the inherent problem of re-tracing the true diffusion path of retweet tree of a diffusion process, we applied an approach previously used before in [insert ref of BMC paper , and [8,9] of the paper]. After re-constructing the diffusion path of the retweet cascades we used the measure of structural virality to quantify the mode of message diffusion, which is the average distance between all pairs of retweeters in the tree. There are multiple measures to calculate the structural virality of a diffusion tree as discussed briefly in the work of [insert ref], and we use their proposed approach of using Weiner's index to compute the structural virality of the diffusion cascade.

d) **Variation in Information Contagion Dynamics**

We investigate the spread of the tweets from the perspective of information adoption, asking the question , " How do successive exposures to a tweet through mutual retweets to a user affect the probability that the user will retweet the information as well ? ". Different from the work of [Romero et al.], we build a network on the users from the structure of interaction via retweet signals , instead of mention signals. We use the retweet metadata and friendship graph available , used in the previous sections above to get the exact timestamp and diffusion paths of the adoption of information through retweets. Borrowing the terminology from [24], we call an user is k-exposed to a seed user *h ,* if he has not retweeted a message from *h ,* but k of his friends have broadcasted message of *h.* We then estimate the probability that user *u* , which has been k-exposed to a user *h* ,will retweet a message from *h* before becoming (k+1) exposed . Let E(k) be the number of users who were k−exposed to h at a certain interval , and let I(k) be the number of users that were k−exposed and retweeted h before becoming (k + 1)−exposed. We then conclude that the probability of retweeting the user h while being k−exposed to h as p(k) = I(k)/ E(k) .The *p(k)* curve averaged over all the retweet chains thus constructed will help in further analyzing the diffusion behavior through characteristic analysis of the *stickiness* and *persistence* of the curve.

e) **Meme Analysis**

We analyzed how the emotional attributes of the content produced by the bots and humans differ from the content they spread or retweet. We re-use the methodology used in *Section B* to annotate the emotional composition of the tweets. Most of the prior works [insert Ref of Ferrara's PeerJ, Dickerson, Ferrara's Social Bots] study the emotional dynamics of the tweets using sentimental analysis tool, SentiStrength [insert ref SentiStrength] to annotate tweets with positive and negative polarity score ranging from 1-5. Similarly, the work of [25] associated tweets with intensity scores from eight basic emotions according to Plutchik's wheel of emotions. In our work, we focus our study towards the conversational aspects of the political communication. A drawback of our work is that we totally eliminate positive dimensions of emotions under study and focus only towards the negative family of emotions. But, we hope to capture the fine resolutions of negative activities of user behavior demonstrated on the tweets to explain the conversations in terms of user attitude. We divide the tweets of the bots and humans into the ones authored by them and the ones retweeted, namely *bots_pure, bots_rt* and *humans_pure,humans_rt*. We perform our analysis on two different levels , by comparing *bots_pure* with *bots_rt* and then expanding the comparison with the humans. We repeat the same analysis for humans. To diversify our study, we divide our study based on spreading of three different types of messages . The first of the types of messages are about conversation of the political candidate, identified by the usage of *#Renzi*. The second subject of conversation we analyze is about the opposition political candidate , identified by the usage of *#beppegrillo .* And our final subject of analysis is about the current news or trending events during the time of election. We thus identify these subsets of messages by the usage of *#Notizie,#Politico140 .* We perform inter group and intra group analysis using different spider plots of the three types of memes under the five different attributes of emotional contents.

f) **Innovators , Adopters, Majorities and Laggards**

We analyzed the process of information diffusion observed in the bots and humans using the idea of Diffusion of Innovations . The theory of Diffusion of Innovations [26] helps us in understanding the spread of an innovation through user motivations and adoption behaviors. The theory proposes five different characteristic segments of users on any possible diffusion processes ; the Innovators , the Early Adopters, the Early Majority, the Late Majority and Laggards. A diagrammatic representation of the user segment, along with the notational percentages of each segment, relative to the timeline of users participating in the election is shown in Figure [insert the Everett's curve .] For each user in our study, we aggregate all of the retweeters through all of his tweets, alongside the difference in timestamp of their retweets since the original tweets , calling the attribute "time elapsed for adoption". We then sort all the users influenced by a user by their "time elapsed for adoption" attribute and divide the users into multiple segments of the diffusion curve. We apply the same process for the bots and human users.For the first part of our analysis , we identify the variance of users in each individual segment of the curve , and take an average across all of the users , to give a notion of uniqueness of innovators reached by the users. For the second part of our study, we compare the variance of the segment membership across different users of the botnets/ human datasets. To perform this analysis, we construct a "Following" network of all the participant retweeters and the source users, then extract the Giant component in each of the cases of humans and bots for further analysis. We performed this further step to make sure that the study of variance of segments across the different users is considerably reasonable in terms of studying a community of users which actually represents strong sharing of innovation.

g) **Longitudinal Analysis of User Roles**

We divide the users present in our datasets into four different roles , previously proposed by [insert Gonzalez-Bailon et al] to determine user roles as a function of their social connectivity and diffusion. We define information diffusion of a user as the # of retweets sent / # of times retweeted , and social connectivity of a user as # followees / # friends. The users are then divided into four roles : influential users (social connectivity <1 and information diffusion < 1,) , hidden influentials (social connectivity > 1 and information diffusion <1) , common users (social connectivity >1 and information diffusion >1) and rebroadcasters (social connectivity <1 and information diffusion > 1).

We divided the tweets occurring in the datasets of human users and bots into multiple time buckets , with each time bucket corresponding to 60 days (2 months) .

Our methodology varies from the work of [insert Social upheaval reference] in that we consider a larger timeline for each time bucket considered to their work (1 day). Our goal is to analyze the longitudinal role change through a fairly stable time period for a social process such as election. We then recomputed the role membership of each user in each of the time bucket , and assigned a user multiple roles across time for the longitudinal analysis. We then study how stable are the user roles across time by studying the volume of user role memberships and the stability of roles across time.

4. Experimental Results

a) Eco-Chambers and their extent

In *Table 2*, we list multiple statistics regarding the production and consumption polarities for the bots and human users. The Pearson correlation between Production polarity and Consumption polarity for the bots under study is -0.0444 compared to the human user's 0.258.The negative correlation observed in bot's sharing activity and it's spreaders is very shocking compared to a relatively stronger Pearson correlation in case of humans, highlighting strong case of eco-chamberism present in the community of human users. On average, the bots produce content with lower polarity score than the humans, and the mean polarity score of its content consumers is also lower than the human users. The Standard Deviation (S.D) of the consumption polarities of the bots is higher for the bots than humans, suggesting the bots were able to reach audiences with relatively broader political spectrums. Secondly, we shift the analysis asking the question on how the increase in production polarity affects the variance/spread of polarity scores of the consumers reached by the producers. In *Figure 1*, we plot the spreader influence variance against the production polarity. It can be seen that in the case of bots, with the polarity scores being very close to zero itself, the spread of polarity scores is very high as compared to the humans. The production polarity of bots are in very lower ranges compared to humans, and they still attract higher ranges of variance in polarity scores of spreaders than the humans. In case of the human users, as the produced content become more polar, the variance of the spreaders reached decreases significantly , again reflecting a very strong presence of eco-chamberism.

Measure	Bots	Humans
Correlation between Production Polarity and Consumption Polarity	**-0.04441325**	0.25828625

Mean Production Polarity	**-0.0021062187452807655**	-0.03877753959710347
Mean Consumption Polarity	**-0.00036488270958366604**	-0.03111847415877329
Std. Production Polarity	**0.005547265326746134**	0.07432802295090955
Std. Consumption Polarity	**0.18765716021810924**	0.12334120725419913

Table 2: Measures of Polarity for Bots and Humans

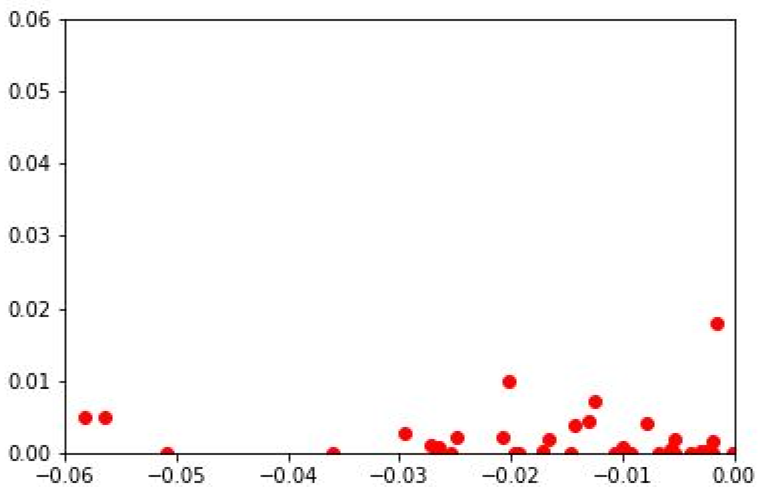

Fig 1(a): Production Polarity vs Spread of Influence in Humans

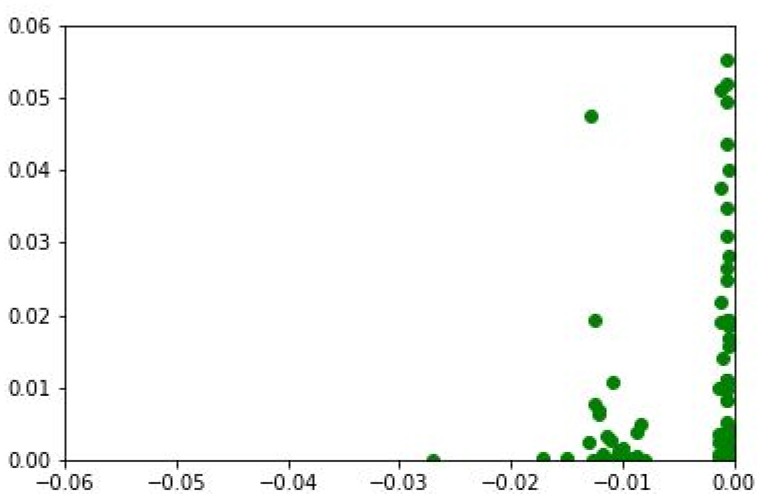

Fig 1(b) : Production Polarity vs Spread of Influence in Bots

b) Emotional Cost of Information Diffusion

We then plot each of the emotional dimensions of the tweets in Y- axis against the influence (retweet) gained by the tweet in X-axis.

We study how the emotional attribute score changes with the increase in the amount of influence gained by the tweets. In *Figure 2*, for all the emotional attributes of the tweets we are studying, we can observe two different types of behaviors : firstly, the emotional scores of the diffused tweets by humans are always higher than those by the bots (both deleted as well as active) , with increasing amount of gathered influence . The bots were able to gain the same amount of influence as humans by tweeting content with significantly lesser amount of toxicity, identity attack, profanity. We also observed that with the increasing amount of influence achieved, the tweets of the deleted bots were slightly in the upper ranges of the emotional score, which can be interpreted as the reason those bots were deleted from the platform, while the one's active are still prevalent. Most interestingly, we can observe that in case of the human users, the increasing influence comes with the cost of increase of emotional scores , for all of the negative emotions under study. Whereas, the bots had only a little or no rise of negative emotion with respect to the increase of influence obtained. This symbolizes that the emotional cost of diffusion was higher for the human users than the bots. The phenomenon of positivity bias in information spread has been validated in prior works thus explaining tweets with positive sentiments spreading more in volume than the negative ones. Our findings suggest that the phenomenon was stronger for the social bots than the human users. Our findings disagree with the one reported by [insert why so emotional paper], who reported that emotionally neutral tweets failed to garner much attention in terms of Retweet. Building upon our findings of the bots tweeting with lower emotionally harmful sentiments than the human users, we also studied how the speed of diffusion correlates with the negative emotional attributes of tweet. We re-use the definition of speed of diffusion used in prior works [27], as the duration between the time the tweet was posted and it's first retweet observed. We observed that the bots were able to gain much lesser response time to the first re-tweets on average (15839.53 seconds) , implying greater speed than the humans (81371.83 seconds), by tweeting with lesser magnitude of all emotional dimensions on their text at the same time.Our findings are contrasting with the work of [27], where the author reported tweets with higher negative valence having higher speed in spreading than the positive and neutral ones.

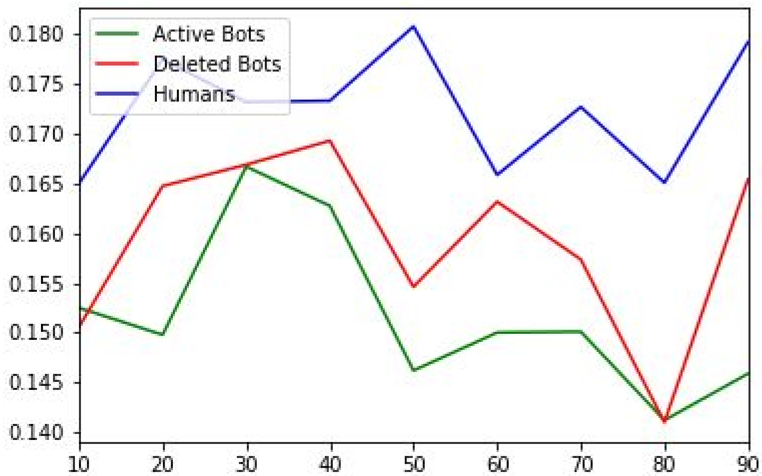

Figure 2(a) : Cost of Diffusion (Severe Toxicity)

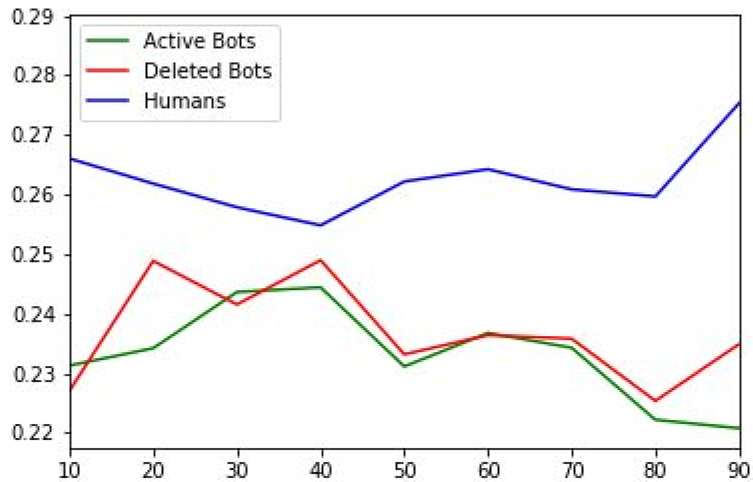

Figure 2(b) : Cost of Diffusion (Threat)

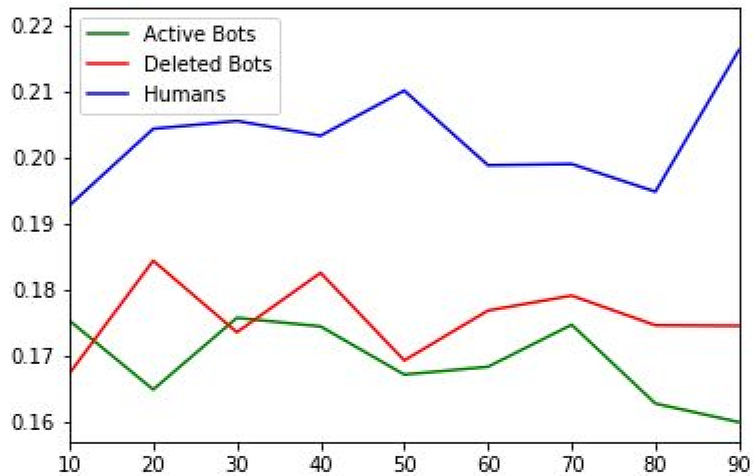

Figure 2(c) : Cost of Diffusion (Identity Attack)

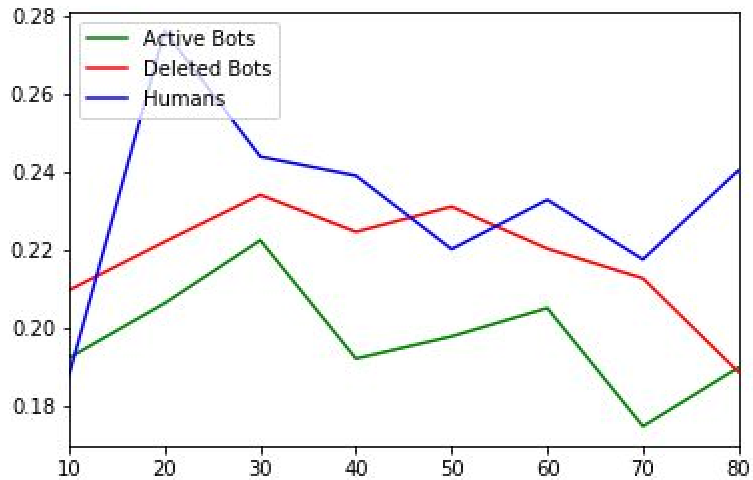

Figure 2(d) : Cost of Diffusion (Obscene)

c) **Mechanisms of Information Spreading**

After reconstructing the diffusion tree using the methodology discussed , we measure the structural virality of information cascades.To get an initial idea about the popularity of the seed messages for our analysis , we plot the distribution of cascade sizes for both bots and humans in *Figure 3*. We can observe that the messages diffused by bots had larger cascade sizes, denoting the popularity of messages authored by bots than humans. We then plotted the structural virality of the diffusion cascades in *Figure 4*. A higher value of structural virality would represent viral mode of diffusion being more dominant in the communication patterns. We can observe that the information cascades of the bots have higher values of structural virality than by humans. This signifies that the spread of information of tweets authored by humans mainly followed a broadcasting pattern, indicating a star network of retweets from the original tweets but without much further retweets and selective sharing with declining diversity over time . Whereas , the higher value of structural virality for the diffusion cascades of the bots in our study signifies a complex process of viral spreading diffusion occurring in case of bots. This signifies the more effective mode of communication pattern occurring in the diffusion process of twitter bots, where messages went “viral” through multiple chains of individual-to-individual diffusion process, also indicating higher probability of cross ideological sharing across heterogeneous communities, which might also help explain the results of section I.

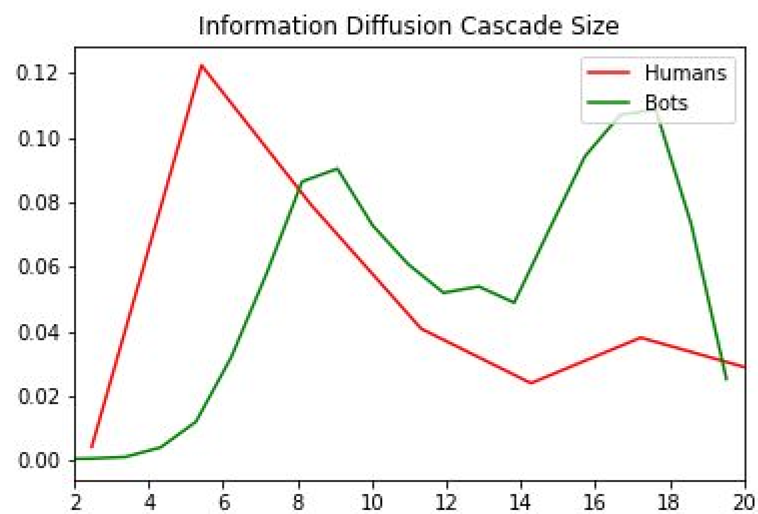

Figure 3 : Information Diffusion Cascade size comparison

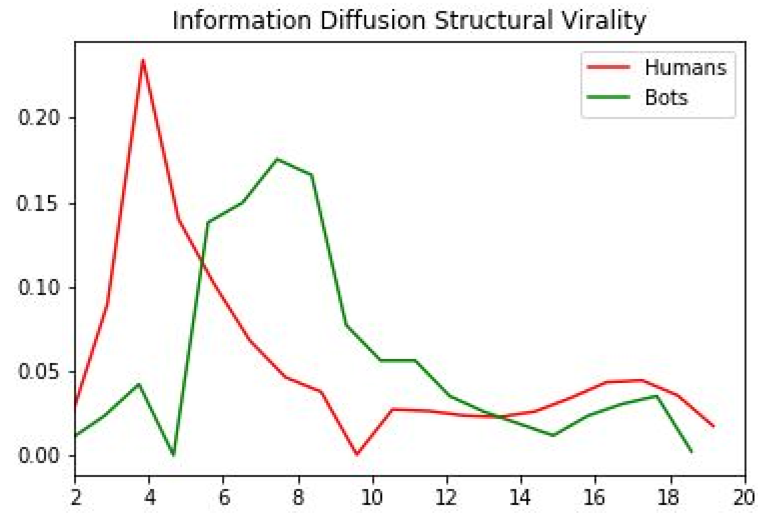

Figure 4 : Information Diffusion Structural Virality comparison

d) **Variation in Information Contagion Dynamics**

We calculated the exposure curves *P(k)* for each of the information diffusion chains of the bots and humans, and averaged those curves curves in *Figure 5*. The stickiness of the information spreading, signified by the maximum value of *P(k)* is considerably higher for the human users than the bots as we can see a ramp up to peak value reached relatively early followed by decline for larger values of *k.* Now, when we compare the persistence behavior, visible through the rate of decay after the peak of the curve,we can observe a significant differences. In the case of humans, the adoption probability drastically falls with increasing magnitude of exposure, indicating less effect of repeated exposure in increasing the adoption by their followers. Interestingly, for the bots, the adoption rate remains constant with increased exposure level from *k :* 1 - 4, and suddenly increases for *k :* 4-6 before falling off eventually. The shape of the curve represents higher persistence of conversation for tweets by the bots compared to the humans. The more sticky and less persistent behavior of the human exposure curve suggests that if the audience of the human users don't adopt or retweet the idea put forward by them after the initially small number of exposures, the chances the idea gets adopted later on are very marginal. The repeated exposures having significant marginal effects on the probability of retweeting the seed user after *kth* exposure hints towards the behavior of complex contagion present in the communication demonstrated by the tweets of the bots.

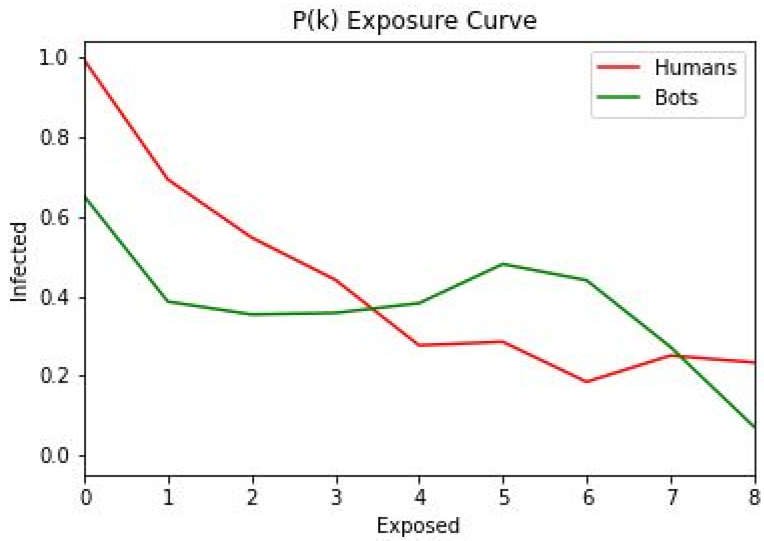

Figure 5 : P(k) Exposure Curve for Bots and Humans

e) **Meme Analysis**

We plot the three radar charts for analyzing the emotional content of the bots and humans about the three different topics of conversations in *Figure 6*. We get very interesting findings from the radar chart. Firstly, the intensity of negative emotions for the tweets authored by the bots is significantly lower than of the tweets they retweet from human users. Whereas, for the human user datasets, it’s exactly the opposite, as the tweets authored by humans have higher intensity of all the negative emotions than the ones they retweet. We also note that the overall intensity of the tweets by bots , whether written by them or forwarded is always smaller than that of the humans. This signifies that the bots were constructing their tweet with significantly less aggressive sentiment than of the tweets they were forwarding.While in the case of humans, they were already putting their own ideas with more negative emotional score on the topics than the content they rebroadcasted , while also tweeting with more negative emotional aggressiveness than the bots under study. The pattern of results were consistent across all of the three types of discussion messages we studied i.e supporting political candidates, opposition political candidates and news discussions. Our findings agree from the conclusions of [1] that the bots produce systematically more contentious (in our case, less toxic) support of a candidate they are advocating for , signifying grassroot support for a given candidate. But, our diverse context of analysis also suggests us that the overall lack of negative aggression is not only with regards to candidate of their interests, but also about the opposition political candidates, and general news based topics. The bots tend to stay away from including multiple forms of toxic elements in their tweets while discussing even the topics and people of the other side of the spectrum, where one could expect them to talk about negatively. The relatively larger presence of toxicity content in the human users with regard to the news based conversations as well as both sides of the political spectrums confirm the findings of [insert Analysis of political disclosure paper] where they reported human users having larger negative average sentiment discussing both sides of the political spectrum.

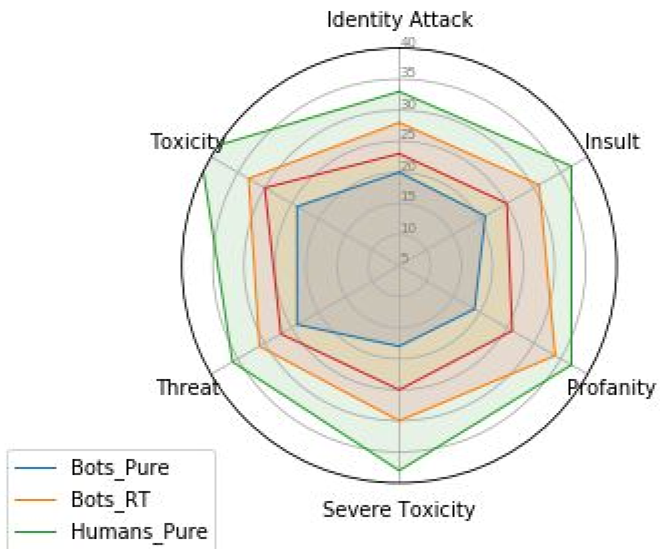

Figure 6(a) : Meme Analysis Spider Chart

f) **Innovators , Adopters, Majorities and Laggards**

We graph the intersection between different regions of information adopters for the bots and human users in *Figure 8* . In our first study, we investigate the spread of information adopters through individual accounts by calculating the proportion of unique adopters in each of the innovation regions and averaging them for all the accounts. We find that the bots have higher proportion of unique users in all of the segments compared to the humans by a significant margin. This hints us that each of the user accounts in the botnets were able to influence a wider variance of users, and even more unique users in the Innovators and Early Adopters region of the innovation cycle. We then repeat the same analysis in a larger scale to study how effective the accounts were, as a community by studying all of the users under the botnet together, and calculating the proportion of unique users in each of the innovation cycle segments. The botnet was able to influence a greater proportion of unique users in each of the innovation cycle than the human users in each of the innovation cycle. Both of our results signifies the bots, individually and as a community influenced larger unique audiences throughout the diffusion process.

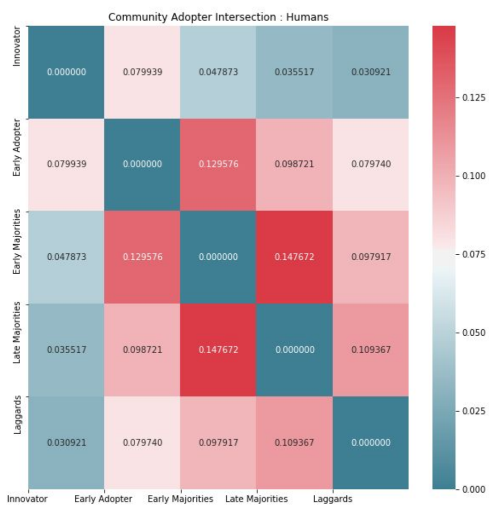

Figure 7(a): Community Adopter Intersection : Bots , Humans

Figure 8(a) : Individual Adopter Intersection : Bots , Humans

g) **Longitudinal Analysis of User Roles**

We plot the longitudinal analysis of user role evolution of humans and bots in our study in *Figure 9*. The proportion of individuals in each class out of the total users at each time frame are represented by the size of the circle. From the figures, we can analyze that the information consumers maintain a considerable proportion of appearance across the timeline for human users, whereas in the bots the activity of the role is very feeble. Interestingly, the bots have continuously strong presence of hidden influentials across the timeline , whereas human users have little to no presence of hidden influentials in any distinct time phase. The most prominent and stable of the roles on human users is of the "Rebroadcasters", while the bots have very less proportion of rebroadcasters occurring at only a few time intervals with smaller proportions. This result is somewhat surprising because the bots are expected to consistently play a role of rebroadcasters based on the objective of their deployment, but we observed the increasing ratio of human users instead of the social bots adapting to that role particularly intriguing. Next, when we analyze the displacement of the individual roles of the bots and human users, the human users are more unstable in the memberships for all of the roles over time. They have more average displacement for every role types, both by the frequency and magnitude of the displacement. Whereas, there is very little fluctuation on the roles of the bots , with smaller magnitude of displacement of user roles. In both of the cases of humans and bots, the majority of the role transitions occurred between Rebroadcasters to Influentials and Influentials to Rebroadcasters.

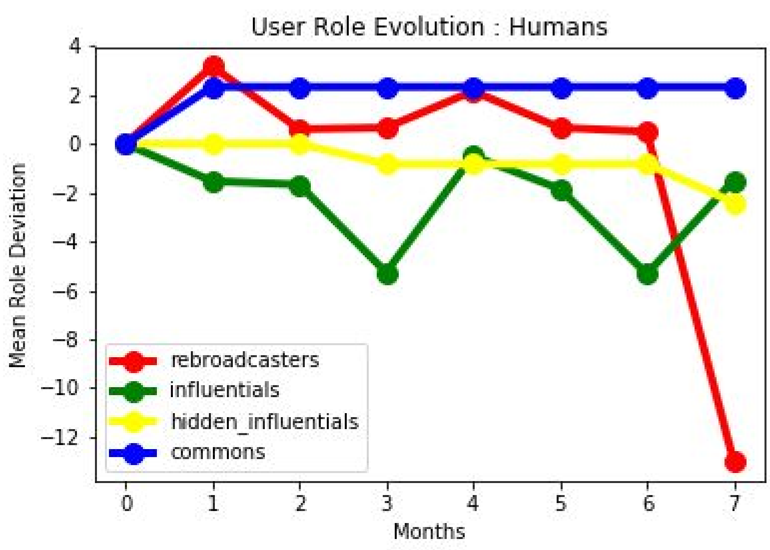

Figure 9(a) : User Role Evolution on Humans

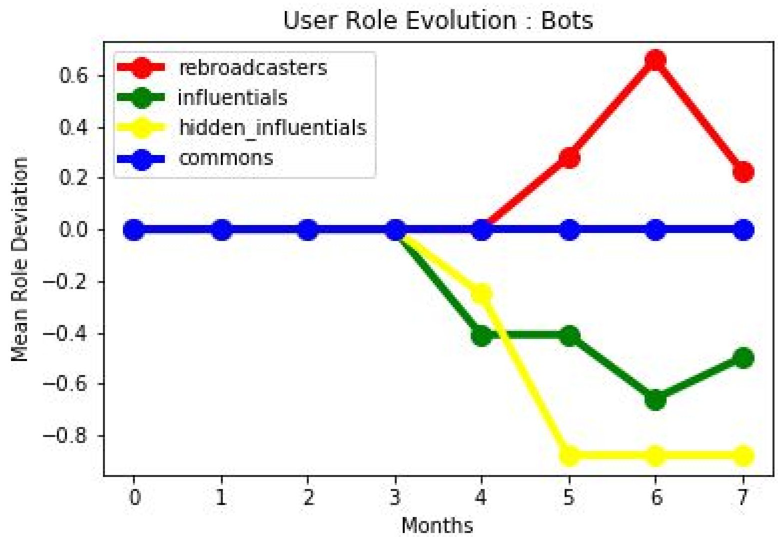

Figure 9(b): User Role Evolution on Bots

References

1. Bessi, Alessandro, and Emilio Ferrara. "Social bots distort the 2016 US Presidential election online discussion." *First Monday* 21.11-7 (2016).
2. Howard, Philip N., and Bence Kollanyi. "Bots,# StrongerIn, and# Brexit: computational propaganda during the UK-EU referendum." *Available at SSRN 2798311* (2016).
3. Forelle, Michelle, et al. "Political bots and the manipulation of public opinion in Venezuela." *arXiv preprint arXiv:1507.07109*(2015).
4. Woolley, Samuel C. "Automating power: Social bot interference in global politics." *First Monday* 21.4 (2016).
5. Abokhodair, Norah, Daisy Yoo, and David W. McDonald. "Dissecting a social botnet: Growth, content and influence in Twitter." *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 2015.
6. Mustafaraj, Eni, and P. Takis Metaxas. "From obscurity to prominence in minutes: Political speech and real-time search." (2010).

7. Hegelich, Simon, and Dietmar Janetzko. "Are social bots on Twitter political actors? Empirical evidence from a Ukrainian social botnet." *Tenth International AAAI Conference on Web and Social Media*. 2016.
8. Keller, Franziska B., et al. "How to manipulate social media: Analyzing political astroturfing using ground truth data from South Korea." *Eleventh International AAAI Conference on Web and Social Media*. 2017.
9. BBC. (2016). EU referendum petition hijacked by bots. Retrieved from http://www.bbc.com/news/ technology-36640459
10. Ferrara, Emilio. "Disinformation and social bot operations in the run up to the 2017 French presidential election." *First Monday* 22.8 (2017).
11. Kollanyi, Bence, and Philip N. Howard. "Junk news and bots during the German parliamentary election: What are German voters sharing over Twitter." (2017).
12. Ratkiewicz, Jacob, et al. "Detecting and tracking political abuse in social media." *Fifth international AAAI conference on weblogs and social media*. 2011.
13. Shao, Chengcheng, et al. "The spread of low-credibility content by social bots." *Nature communications* 9.1 (2018): 4787.
14. Rathnayake, Chamil, and Wayne Buente. "Incidental effects of automated retweeting: An exploratory network perspective on bot activity during Sri Lanka's presidential election in 2015." *Bulletin of Science, Technology & Society* 37.1 (2017): 57-65.
15. Woolley, S., and Phil Howard. "Bots unite to automate the presidential election. Wired." (2016).
16. Everett, Richard M., Jason RC Nurse, and Arnau Erola. "The anatomy of online deception: What makes automated text convincing?." *Proceedings of the 31st Annual ACM symposium on applied computing*. ACM, 2016.
17. Zhang, Jinxue, et al. "On the impact of social botnets for spam distribution and digital-influence manipulation." *2013 IEEE*

Conference on Communications and Network Security (CNS). IEEE, 2013.

18. Woolley, Samuel C., and Philip N. Howard. "Automation, algorithms, and politics| political communication, computational propaganda, and autonomous agents—Introduction." *International Journal of Communication* 10 (2016): 9.
19. Paudel, Pujan, et al. "How the Tables Have Turned: Studying the New Wave of Social Bots on Twitter Using Complex Network Analysis Techniques." (2019).
20. Cresci, Stefano, et al. "The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race." *Proceedings of the 26th international conference on World Wide Web companion*. International World Wide Web Conferences Steering Committee, 2017.
21. Llewellyn, Clare, Laura Cram, and Adrian Favero. "Avoiding the Drunkard's search: Investigating collection strategies for building a Twitter dataset." 2016 IEEE/ACM Joint Conference on Digital Libraries (JCDL). IEEE, 2016.
22. Henrique, J. "GetOldTweets-python [Computer software]." (2016).
23. "https://www.perspectiveapi.com/"
24. Romero, Daniel M., Brendan Meeder, and Jon Kleinberg. "Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter." *Proceedings of the 20th international conference on World wide web*. ACM, 2011.
25. Kušen, Ema, and Mark Strembeck. "Why so emotional? An analysis of emotional bot-generated content on Twitter." *Proc. of the 3rd International Conference on Complexity, Future Information Systems and Risk (COMPLEXIS)*. SciTePress, 2018.
26. Rogers, Everett M. *Diffusion of innovations*. Simon and Schuster, 2010.

27. Ferrara, Emilio, and Zeyao Yang. "Quantifying the effect of sentiment on information diffusion in social media." *PeerJ Computer Science* 1 (2015): e26.
28. “https://mediabiasfactcheck.com/”