The Fake News Spreading Plague: Was it Preventable?

Eni Mustafaraj
Department of Computer Science
Wellesley College
Wellesley, MA
eni.mustafaraj@wellesley.edu

Panagiotis Takis Metaxas
Department of Computer Science
Wellesley College
Wellesley, MA
pmetaxas@wellesley.edu

ABSTRACT

In 2010, a paper entitled "From Obscurity to Prominence in Minutes: Political Speech and Real-time search" [7] won the Best Paper Prize of the WebSci'10 conference. Among its findings were the discovery and documentation of what was labeled a "Twitter bomb", an organized effort to spread misinformation about the democratic candidate Martha Coakley through anonymous Twitter accounts. In this paper, after summarizing the details of that event, we outline the recipe of how social networks are used to spread misinformation. One of the most important steps in such a recipe is the "infiltration" of a community of users who are already engaged in conversations about a topic, to use them as organic spreaders of misinformation in their extended subnetworks. Then, we take this misinformation spreading recipe and indicate how it was successfully used to spread fake news during the 2016 U.S. Presidential Election. The main differences between the scenarios are the use of Facebook instead of Twitter, and the respective motivations (in 2010: political influence; in 2016: financial benefit through online advertising). After situating these events in the broader context of exploiting the Web, we seize this opportunity to address limitations of the reach of research findings and to start a conversation about how communities of researchers can increase their impact on real-world societal issues.

CCS CONCEPTS

• **Information systems** → **Social networks**; *Spam detection*; • **Human-centered computing** → **Social networking sites**;

KEYWORDS

fake news; misinformation spreading; Facebook; Twitter; Google

ACM Reference format:
Eni Mustafaraj and Panagiotis Takis Metaxas. 2017. The Fake News Spreading Plague: Was it Preventable?. In *Proceedings of ACM Web Science Conference, Troy, NY, USA, June 2017 (WebSci'17),* 5 pages.
https://doi.org/10.1145/3091478.3091523

WebSci'17, June 2017, Troy, NY, USA

ACM ISBN 978-1-4503-4896-6/17/06…$15.00
https://doi.org/10.1145/3091478.3091523

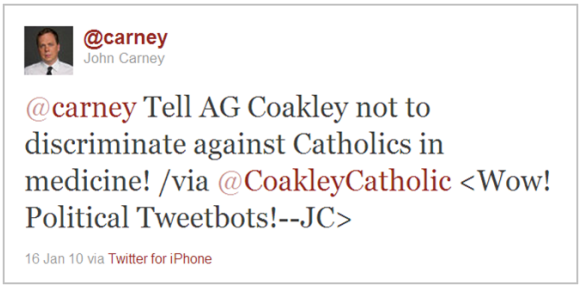

Figure 1: The journalist John Carney (at that time with CNBC), received one of these "reply-tweets", which he retweeted adding a comment expressing his surprise, "<Wow! Political Tweetbots!–JC>", because this was an unknown phenomenon at that time on Twitter. Carney deleted the URL of the original tweet.

1 INTRODUCTION

1.1 The Anatomy of a Political Twitter Bomb

On January 15, 2010, between 18:43 and 18:56, someone created a group of nine Twitter accounts with the names: *CoakleySaidWhat*, *CoakleyCatholic*, *CoakleyER*, *CoakleyAgainstU*, *CoakleyAG*, *CoakleyMass*, *CoakleyAndU*, *CoakleyWhat*, and *CoakleySaidThat*. The name Coakley refers to Martha Coakley, at that time the Attorney General of Massachusetts and the democratic candidate running in the high-stakes Special Election for the Massachusetts U.S senate seat. After a few hours of inactivity, these nine accounts sent 929 tweets addressed to 573 unique users in the course of 138 minutes. All the tweets contained a URL to the same website `http://coakleysaidit.com`, (also registered on January 15, 2010), that showed video and audio from a speech by Martha Coakley, taken out of context, to advance the false claim that she is against the employment of Catholics in the emergency room.

The nine accounts were sending a tweet per minute and repeating more or less the same content, both reasons to be flagged as a spamming account. Twitter discovered the automated tweets and consequently suspended all nine accounts. Their existence and their misinformation attack would have gone unnoticed had it not been for one fortunate circumstance: we were collecting all tweets containing the names "coakley" and "brown" (respectively for Martha Coakley and Scott Brown, the two candidates for the senate election) in real-time, during the week leading to the election. The tweets sent by these anonymous accounts were no simple tweets, they were so-called "reply tweets", tweets directed to particular users. Why? Because a new account on Twitter doesn't have any followers. Tweets sent by such an account will not be read by anyone. Thus, directing the tweets to a particular user makes it likely

that the tweet will be read. But to which users do you reply-tweet, out of the millions that are on Twitter? This is where a common spamming technique on Twitter comes in handy: reply to users who have used certain desired keywords in their tweets, that is, to users already attuned to the topic. Our analysis of the recipients of these "reply tweets" revealed that 96% of them had been tweeting about the MA senate race in the four-hour interval between the time the anonymous accounts were created and when they started to send the "reply tweets". Almost 25% of the users who received a tweet (143 out of 573) retweeted the message. A screenshot from one of the retweets is shown in Figure 1. We chose to show this tweet, because the user is a well known journalist[1] and an experienced Twitter user who joined the site on March 2007. His surprise at the message indicates the novelty of this technique at the time. The retweets had the effect that the followers of the retweeters were likely exposed to the misinformation, which they would have not seen otherwise, given that the messages didn't include hashtags, a common way to group together tweets about a topic, which, when clicked, present with a stream of tweets containing the hashtag. Our estimation of the audience size, based on the followers of the retweeters, amounted to 61,732 Twitter users.

1.2 A Recipe for Spreading Misinformation on Twitter

All the facts presented in the previous subsection were part of the WebSci'10 paper [7]. What we didn't do in that paper was to summarize our findings in an easily memorable recipe, which contains the steps used by the propagandists in spreading their misinformation on Twitter. We're providing this recipe for the first time in this paper.

Step 1	Register a domain name for a new website, for example: `http://coakleysaidit.com`
Step 2	Create anonymous accounts, for example: *CoakleySaidWhat*, etc.
Step 3	Identify a community of users interested in the topic, for example, the MA Senate Election race.
Step 4	Target members of this community with messages, for example, reply to users providing link to website.
Step 5	Wait for members of community to spread message via retweets in their organic subnetworks.

Table 1: A recipe for spreading misinformation on Twitter via a Twitter bomb.

Our discovery attracted the attention of both journalists and other researchers. A team at Indiana University, headed by Fil Menczer, developed Truthy[2], a system that collects Twitter data to analyze discourse in near real-time [12]. In addition, our team at Wellesley developed Twitter Trails[3], a system that can be used to monitor the spreading of rumors on Twitter [6]. This focus on Twitter is justified by the fact that it provides APIs for researchers to collect and analyze its data, as well as the public nature of conversations on Twitter. Both these features are missing on Facebook (not entirely, but they are severely limited), thus, only Facebook employees are able to study them. As evidence, see [1]. Meanwhile, researchers not affiliated with the company have almost no opportunities to study information spreading on Facebook, especially that of rumors, hoaxes, and recently fake news, a topic to which we turn our focus now.

1.3 Spreading Fake News on Facebook

After the surprise results of the 2016 U.S. Presidential Election, the American media directed its ire at Facebook and Google, as in this New York Times piece [19] written by the Editorial Board, on November 19, 2016:

> Most of the fake news stories are produced by scammers looking to make a quick buck. The vast majority of them take far-right positions. But a big part of the responsibility for this scourge rests with internet companies like Facebook and Google, which have made it possible for fake news to be shared nearly instantly with millions of users and have been slow to block it from their sites.

This criticism is only partly well-placed. Facebook had been working toward fixing (or containing) the spread of hoaxes on the site at least since January 2015, almost two years before the election [11]. They defined a hoax as a form of News Feed spam post that includes scams ("Click here to win a lifetime supply of coffee"), or deliberately false or misleading news stories ("Man sees dinosaur on hike in Utah"). As we can notice from this definition, in 2015, the phrase *fake news* wasn't being applied yet to the kind of false stories that flooded Facebook in the weeks before the election.

Step 1	Register many web domains for related websites, with catchy names such as: `http://TrumpVision365.com`, see [16].
Step 2	Create Facebook accounts of fictitious people, e.g, Elena Nikolov or Antonio Markoski, see [17].
Step 3	Identify and join a Facebook group about a political candidate, e.g., "Hispanics for Trump" or "San Diego Bernicrats", see [17].
Step 4	Target members of the Facebook group with posts that link to the fake news website stories, see [17].
Step 5	Wait for members of the group to spread the fake news in their organic subnetworks, by sharing and liking it.

Table 2: The recipe for spreading *fake news* on Facebook ahead of the 2016 U.S. Presidential election. It contains the same steps as the recipe shown in Table 1.

However, since it was difficult for independent researchers to know the extent to which Facebook users were affected by this issue, everything continued more or less as before, and Facebook was alone in its fight. This changed in early 2016, when the online

[1] John Carney, https://twitter.com/carney.
[2] Truthy, now known as OSoMe, http://truthy.indiana.edu
[3] TwitterTrails.com, http://twittertrails.com

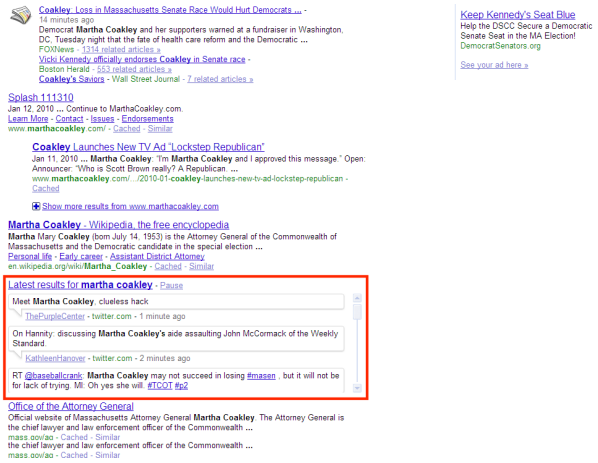

Figure 2: Screenshot from Google search results about Martha Coakley on Jan 12, 2010. Notice in highlighted red, the tweets attacking Coakley. This was a finding from our WebSci'10 paper, on how Google was inadvertently giving premium space to political propagandists, in an effort to have "fresh" and relevant search results.

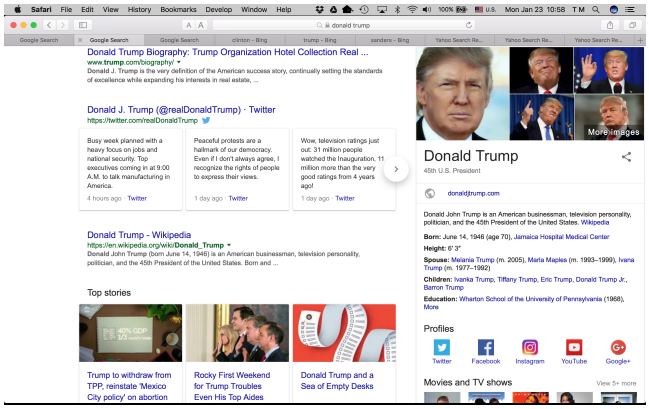

Figure 3: Screenshot from Google search results about Donald Trump on Jan 23, 2017. In addition to the many sections on the page (such as the "featured snippet" on the right column), notice how the tweets shown above the fold belong to Trump himself.

publication BuzzFeed took an interest on Facebook's unsuccessful efforts to deal with the problem. In an article published in April 2016, BuzzFeed proclaims: "it is the golden age of fake news" [4]. The article reveals that BuzzFeed—using the services of the company *Crowdtangle*, specialized in measuring social engagement—had conducted a study of fake news that was spreading via nine known fake news sites, such as the *National Report*, *Huzlers*, or *Empire News*. The findings emphasized that while traffic for these sites had gone down for a while during 2015, it had started picking up again in early 2016. The article also interviewed Allen Montgomery, a fake identity for Jestin Coler, the creator of a factory of fake news websites, as NPR reporters discovered after the election [18]. Coler's interview sheds light about some of the tricks of the trade of fake news, and points out why he believes he can win over Facebook:

> Coler believes Facebook is fighting a losing battle, doomed to fail no matter what it does. "They can shut down *National Report*. Of course they can do that," he said. "But I could have 100 domains set up in a week, and are they going to stop every one of those? Are they now going to read content from every site and determine which ones are true and which ones are selling you a lie? I don't see that happening. That's not the way that the internet works."

Despite this sounding of alarm bells by BuzzFeed (as early as April 2016), things got only worse with fake news on Facebook. We counted at least 25 articles published on the topic of fake news from April to November 2016 on BuzzFeed, culminating with the story of "How teens in the Balkans are duping Trump supporters with fake news", published on November 3, 2016 and followed up by the related piece on "How Macedonian spammers are using Facebook groups to feed you fake news". These two articles provide details about one of the fake news factories operated by young people in the small town of Ceres, Macedonia, that targeted Facebook users in the United States. After reading these news articles (and others on the same topic), we noticed the clear similarities to the process that lead to the Twitter bomb against Martha Coakley in 2010. In fact, we are able to map the steps in the two recipes one to one, as shown in Table 2. This similarity should not be surprising. Once a spamming technique has been proven successful, it is easily replicated, since the knowledge about its working is also shared on the internet. What should surprise and worry us is the fact that researchers and web platform developers also know about such techniques, but they do little to warn and educate the public of the consequences. It is also unfortunate that tech companies who have been exploited to enable misinformation spreading, do not act proactively or effectively in stopping it. As an example of ineffective action, we discuss in the next section the way Facebook handled the accusation that its news verification was not balanced.

2 FROM PROPAGANDA TO FAKE NEWS

We should not give the reader the impression that **online propaganda** started with Twitter bombs or Facebook fake news. In fact, it is much older than social media, it is as old as the Web itself. However, before the development of search engines it was not easy *for propaganda to find you*. Search engines made it possible for propagandists to spread their message using techniques we now call *Web Spam* [5]. Advertisers, political activists and religious zealots of all kinds have been busy modifying the structure of the Web in an effort to promote their own biased results over organic, unbiased results. The Search Engine Optimization industry has grown out of this effort, sometimes using unethical techniques to promote their messages, and search engines have been continuously evolving to fend off these attacks.

In much of the first decade of the new millennium, search engines tried to defend against Web spam, but the spammers were successful in circumventing their defenses by using first "Link Farms" and

later "Mutual Admiration Societies", collections of web sites that would intentionally link to each other in order to increase everyone's PageRank [5]. Even when Google was reportedly using up to 200 different signals to measure quality [13], professional SEOs would manage to get sites like JC Penney's at the top of search results [14]. Google's ingenious solution to the problem of "unfair competition for high placement on a page" was the introduction of the advertising model of AdWords and AdSense that gave spammers an opportunity to make money while following rules. That seemed to work for a while. However, the algorithmically allocated financial benefits of online advertising became so lucrative, they provided a strong reason for *anyone* to have a presence on the Web—especially if they could manage to attract clicks and thus, advertising dollars. This led to "click bait" and to the creation of ads masquerading as outrageous (fake) news stories, as we discussed in the previous section.

But as search engines and social media evolve, so do the propagandistic techniques. Concurrently with the rise of "fake news", we also find the "featured snippets" manipulation [3], and "autocompletion revelations" [2], as the latest chapters in spreading propaganda through search engines and social media so that it will find you. As a community of researchers, we need to embrace the challenge of documenting and understanding these phenomena, as well as finding effective ways to make these issues known to the platform providers. Journalists also need to be informed, as they sometimes give credence to conspiracy theories, confusing the public [20].

3 RESEARCH THAT INFORMS DESIGN

It is important for researchers, journalists and web users to pay attention continuously to the information and misinformation they encounter on the Web, be it on Google, Twitter, or Facebook. In this section, we discuss how research results and their publicizing lead over time to changes in the design features of these systems, addressing the exhibited weaknesses.

3.1 The Evolution of Google Search Results

The central finding that give the title to our WebSci'10 paper [7] was the manipulation of Google real-time search results through repetition of Twitter posts by accounts—real users or bots—supporting a particular candidate. In December 2009, just one month before the Massachusetts special election for the U.S. Senate seat in 2010, Google followed Bing in introducing "real-time search results", placing social media messages near the top of search results for a relevant search query. These messages came mostly from Twitter, since its API makes it easy to pull the tweets. Tweets appearing in the search results were those that had been recently posted. That created the opportunity for political propagandists to exploit the search results, creating a Twitter-enabled Google bomb. As we documented in our paper, the manipulators were repeating the same messages, something also allowed by Twitter, over and over to increase the volume of conversation about a particular topic and keep it fresh for search engines to include in their real-time results. Repetition of a message would be annoying to the propagandist's followers, but the target was not their followers' feed; it was Google and Bing's algorithms.

We can see these highly-placed tweet messages from random Twitter accounts in the screenshot that we took in January 2010 for Martha Coakley's search results, Figure 2. During 2010, Google eventually recognized that giving anonymous social media accounts a premium spot in its search results was not in line with its goals for reliable information and for a few years this feature disappeared. However, it has come back again, but now in a different format: when searching for a person, it will pull up tweets from their timeline, as opposed to tweets about them, as exemplified in Figure 3. This is a great improvement, because it prevents actors—who have an interest in promoting their adversarial messages about an individual or product—to receive an unearned spot at the top of the search results.

3.2 The Evolution of Retweeting

In [7], we had included the following observation at the end of Section 4:

> Our experiments with Google real-time search have shown that, even though Google doesn't display tweets from users that have a spammer signature, it does display tweets from non-suspected users, even when these are retweets coming from spammers. Thus, simply suspending spamming accounts is not sufficient. There should be a mechanism that allows for retroactively deleting retweets of spam and some mechanism that labels some Twitter users as enablers of spam.

At that time (in 2010), Twitter didn't have an easy way to quote a tweet and it allowed users to edit the original tweet text when retweeting, as the tweet shown in Figure 1 indicated. That design feature turned out to be very problematic, among others for the reason mentioned in the quote above: deleted spam tweets lived in the retweets of other Twitter users, but also because users were often purposefully changing the meaning of the text they were retweeting [9]. Most of this was possible via third-party applications that were very popular in the early years of Twitter. These applications were shut down over the years and nowadays Twitter doesn't allow the editing of a tweet that is being retweeted. Additionally, if the original is deleted, the retweet is deleted too, while in a quoted retweet, the text "This tweet is unavailable." will show in place of the deleted tweet.

3.3 The Evolution of Fake News on Facebook

The proliferation of fake news on Facebook achieved new levels once Facebook made a big change in how its algorithm for the Trending News feature worked. Before August 2016 (when this change took effect), the Trending News feature was being curated by human editors, who filtered out unreliable sources and chose neutral or balanced sources for breaking news stories. However, when the online tech blog Gizmodo posted an article (May 2016) [10], in which former employees of the Trending News lamented anti-conservative bias, Facebook—likely worried about potential lawsuits for suppressing freedom of speech—fired its team of human editors and replaced them with algorithms. It didn't take long after that change for fake news to start achieving Trending News status,

Figure 4: Facebook recently moved into implementing a system of warning towards sharing news items that have been disputed.

as BuzzFeed reported on August 30, 2016 [15]. Despite BuzzFeed's relentless reporting on the fake news plague throughout the pre-election season, the rest of the media and the public didn't tune in into this conversation until after the election.

Facebook initially disputed the it had a fake news problem, claiming that it accounts for only 1% of the news stories. However, the company changed course under the increased and sustained public pressure, introducing new features in its interface and algorithms to address the issue [8].

One important feature that has rolled out recently is the labeling of news posts as "Disputed" via fact-checking, third-party providers such as Snopes or PolitiFact. The screenshot in Figure 4 is an example of this feature in action. In addition to adding this label, Facebook warns users with an alert box before they try to share a disputed story, although they are still allowed to share it [8].

It remains to be seen how this new feature will affect fake news spreading. It bears repeating that the lack of access to Facebook data, which could allow independent researchers to evaluate the effectiveness of such interventions, will hinder our understanding of changes in human behavior correlated with or caused by changes in the socio-technical platforms they inhabit. This is a reason for concern for our research communities.

4 DISCUSSION

What is the moral of the story? In the past, researchers were the ones discovering and documenting the misuse and abuse of socio-technical platforms by the hands of dubious actors with dubious agendas. The WebSci'10 [7] paper is only one such example. However, that discovery was possible only because we were collecting data in real-time, after having noticed some evidence of foul play. When one contemplates Twitter's approach to combating spammers, it seems reasonable that tweets created by "spamming" accounts are automatically deleted and retracted from the entire network, once the accounts are suspended. However, the downside of such an approach is that it makes it impossible for researchers and fact-checkers to go back in time and study the origin of misinformation campaigns and the mechanisms for spreading them. That is a severe limitation to research. The problem becomes even more pronounced in the content of fake news spreading on Facebook. Most Facebook groups are private and if they are the source for starting certain cascades of fake news spreading, outside researchers cannot observe them in their early stages, missing crucial information that would lead to their understanding. Thus, it is not surprising that in the current situation created by the fake news plague, researchers didn't play a leading role in their discovery. It were journalists and not researchers in academia or Facebook and Google who raised concerns, but were not heard. This is worrisome. Facebook, by replacing humans with algorithms, might have played a crucial role in fueling the fake news spreading phenomenon. Similarly, the ease with which Google enables earning ad money for page impression provided the financial incentives for the creation of the fake news industry.

In light of what we know so far, here is our open question to the relevant research communities:

> in the current context of the omnipresent, web-based, socio-technical systems such as Facebook, Google, and Twitter, what decisions should be made by humans and what by algorithms?

Our research communities should lead the way in providing answers to this question.

REFERENCES

[1] Adrien Friggeri, Lada Adamic, Dean Eckles, and Justin Cheng. 2014. Rumor Cascades. In *ICWSM*.

[2] Quentin Hardy and Matt Richtel. 2012-11-21. Don't Ask? Internet Still Tells. *http://nyti.ms/2nyWnSx* (2012-11-21).

[3] Adrienne Jeffries. 2017-03-05. Google's Featured Snippets are worse than fake news. *http://bit.ly/2n5vgB2* (2017-03-05).

[4] Alex Kantrowitz. 2015-01-20. Facebook Wanted A Fight Against Fake News. It Got One. *http://bit.ly/2mBUnIF* (2015-01-20).

[5] Panagiotis Metaxas. 2010. Web Spam, Social Propaganda and the Evolution of Search Engine Rankings. *Lecture Notes BIP, Springer-Verlag* (2010). http://bit.ly/ffYsuC

[6] Panagiotis Takis Metaxas, Samantha Finn, and Eni Mustafaraj. 2015. Using TwitterTrails.Com to Investigate Rumor Propagation *(CSCW'15 Companion)*. 69–72. https://doi.org/10.1145/2685553.2702691

[7] Panagiotis T. Metaxas and Eni Mustafaraj. From Obscurity to Prominence in Minutes: Political Speech and Real-Time Search. In *Proc. of the WebSci'10 Conference.*

[8] Adam Mosseri. 2016-12-15. News Feed FYI: Addressing Hoaxes and Fake News. *http://bit.ly/2mzv8pe* (2016-12-15).

[9] Eni Mustafaraj and Panagiotis Takis Metaxas. 2011. What Edited Retweets Reveal about Online Political Discourse.. In *AAAI Workshop on Analyzing Microtext.*

[10] Michael Nunez. 2016-05-09. Former Facebook Workers: We Routinely Suppressed Conservative News. *http://bit.ly/2ngF2k5* (2016-05-09).

[11] Erich Owens and Udi Weinsberg. 2015-01-20. News Feed FYI: Showing Fewer Hoaxes. *http://bit.ly/2mBUnIF* (2015-01-20).

[12] J. Ratkiewicz, M. Conover, M. Meiss, B. Goncalves, S. Patil, A. Flammini, and F. Menczer. 2011. Truthy: Mapping the Spread of Astroturf in Microblog Streams.. In *WWW'11.*

[13] Saul Hansell 2007. Google Keeps Tweaking Its Search Engine. (2007). URL: http://nyti.ms/2n5JjGx [accessed: 2007-06-03].

[14] David Segal. 2012-02-12. The dirty little secrets of search. *New York Times, http://nyti.ms/2nXfSE0* (2012-02-12).

[15] Craig Silverman. 2016-08-30. Facebook Must Either Innovate Or Admit Defeat At The Hands Of Fake News Hoaxsters. *http://bzfd.it/2nMvfia* (2016-08-30).

[16] Craig Silverman and Lawrence Alexander. 2016-11-03. How Teens In The Balkans Are Duping Trump Supporters With Fake News. *http://bzfd.it/2mC6tBm* (2016-11-03).

[17] Craig Silverman and Lawrence Alexander. 2016-11-08. How Macedonian Spammers Are Using Facebook Groups To Feed You Fake News. *http://bzfd.it/2mzvCM0* (2016-11-08).

[18] Laura Sydell. We Tracked Down A Fake-News Creator In The Suburbs. Here's What We Learned. *http://n.pr/2nuHN1T* (????).

[19] The Editorial Board. 2016-11-19. Facebook and the Digital Virus Called Fake News. *New York Times* (2016-11-19).

[20] Craig Timberg. 2017-03-14. Could Google rankings skew an election? New group aims to find out. *Washington Post, http://wapo.st/2mNa0wY* (2017-03-14).