Computers in Human Behavior 84 (2018) 103–113

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

A computational approach for examining the roots and spreading patterns of fake news: Evolution tree analysis

S. Mo Jang, PhD [a, *], Tieming Geng [b], Jo-Yun Queenie Li [a], Ruofan Xia [b], Chin-Tser Huang, PhD [b], Hwalbin Kim, PhD [c], Jijun Tang, PhD [b, d]

[a] *School of Journalism and Mass Communications, University of South Carolina, USA*
[b] *Department of Computer Science and Engineering, University of South Carolina, USA*
[c] *Healthcare Media Research Institute, Hallym University, South Korea*
[d] *School of Computer Science and Technology, Tianjin University, China*

ARTICLE INFO

Article history:
Received 25 November 2017
Received in revised form
15 February 2018
Accepted 20 February 2018
Available online 22 February 2018

Keywords:
Fake news
Evolution tree analysis
Computational social science
Misinformation
Network analysis

ABSTRACT

To improve the flow of quality information and combat fake news on social media, it is essential to identify the origins and evolution patterns of false information. However, scholarship dedicated to this area is lacking. Using a recent development in the field of computational network science (i.e., evolution tree analysis), this study examined this issue in the context of the 2016 US presidential election. By retrieving 307,738 tweets about 30 fake and 30 real news stories, we examined the root content, producers of original source, and evolution patterns. The findings revealed that root tweets about fake news were mostly generated by accounts from ordinary users, but they often included a link to non-credible news websites. Additionally, we observed significant differences between real and fake news stories in terms of evolution patterns. In our evolution tree analysis, tweets about real news showed wider breadth and shorter depth than tweets about fake news. The results also indicated that tweets about real news spread widely and quickly, but tweets about fake news underwent a greater number of modifications in content over the spreading process.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

The current information environment, notably characterized by the rapid proliferation of social media and partisan outlets, has sparked concerns about the deleterious consequences of fake news and other forms of inaccurate information to democratic societies (Mele et al., 2017; Mihailidis & Viotty, 2017). This concern has reached a climax during the US presidential election in 2016. A number of commentators suggested that the election results were altered by the substantial rise of so-called “fake news” — fabricated stories that were promoted on social media to deceive the public for ideological and/or financial gain (Allcott & Gentzkow, 2017; Jin et al., 2017; Pennycook, Cannon, & Rand, 2017).

One important effort to minimize the impact of fake news, along with other measures such as literacy education, bot control, and fact-checking, is to track down the root source of fake information and to reduce the promotion of such information from those sources (Mele et al., 2017). Thus far, a handful of websites have been reported to be major fake news producers during the 2016 US election period, and researchers have identified more than 100 websites that create and promote fake stories regularly (Shao, Ciampaglia, Flammini, & Menczer, 2016). Others also expressed concerns regarding bots and cyborgs, which may be used to facilitate the spread of false information on digital platforms (Ferrara, Varol, Davis, Menczer, & Flammini, 2016). These efforts and concerns revolve around the source of fake news stories. Therefore, to combat fake news, it is essential to identify the origins and spreading patterns of fake stories on social media (Jin et al., 2017; Mele et al., 2017). If the spread of information from those identified origins can be regulated by algorithms, then the flow of fake news might be mitigated considerably.

Unfortunately, there has been a dearth of scholarship in this area (Goel, Anderson, Hofman, & Watts, 2015; González-Bailón & Wang, 2016). Most efforts using network analysis have focused on the

* Corresponding author.
E-mail addresses: mo7788@gmail.com (S.M. Jang), tgeng@email.sc.edu (T. Geng), joyun@email.sc.edu (J.-Y. Queenie Li), sxia@email.sc.edu (R. Xia), huangct@cse.sc.edu (C.-T. Huang), ku95bini@gmail.com (H. Kim), jtang@cse.sc.edu (J. Tang).

https://doi.org/10.1016/j.chb.2018.02.032
0747-5632/© 2018 Elsevier Ltd. All rights reserved.

identification of influential actors, offering little understanding of the origins of false information. These social network approaches have provided a snapshot of the network density but neglected critical elements for understanding the information-spreading process, such as origin, time, hierarchy, and content evolution.

This study aims to address these issues in the context of fake news tweets during the 2016 US presidential election. Using a recent development in the field of network science (i.e., evolution tree analysis), we examined the root content of fake information, producers of original content, and evolution patterns of fake news tweets. By retrieving 307,738 tweets about 30 fake and 30 real news stories, we also made significant comparisons between fake and real news content as well as between fake news favoring Hillary Clinton and fake news favoring Donald Trump. Our analysis will help elaborate on the following questions: Who is producing the root content of the fake information? Are those producers members of political/financial organizations, journalists, ordinary users, or bots? How do fake and real news stories evolve differently? How does the evolution of fake news differ depending on the partisan orientation of media content (favoring Clinton vs. Trump)? We will use these questions to articulate the evolution of fake and real news stories on social media platforms.

2. Literature review

2.1. Fake news on social media

2.1.1. The rise of fake news

Fake news is defined as misinformation that was fabricated and spread on social media to mislead the audience for political and/or financial gains (Pennycook et al., 2017). Recently, the influx of fake news was thrust into the spotlight during the 2016 US presidential election. Although the exact impact of fake news on the outcome of the election remains unclear, fake news caused some confusion and deception (Pew Research Center, 2016b). Political misinformation is spreading even more widely than before, due to the emerging media environment that streamlines partisan polarization and fragmentation (Bakshy, Messing, & Adamic, 2015). This information ecosystem has spurred more doubt about individuals' "honesty, decency, fairness, patriotism, and sometimes sanity," resulting in fears and anger among the public and an unprecedentedly polarized country (Sunstein, 2009, p. 85).

2.1.2. Combating fake news

A number of pathways for combating fake news have emerged (Mele et al., 2017). First, the literacy approach suggests that education or literacy intervention should be implemented to heighten individuals' cognitive ability to discern fact from fiction (Allcott & Gentzkow, 2017; Jang & Kim, 2018). However, this view of media literacy as a panacea for the dissemination of fake news has received criticism for oversimplifying the problem by shifting responsibility to ill-informed citizens and justifying the current mushrooming of misinformation online (Mihailidis & Viotty, 2017). The second pathway suggests a corrective approach by making fact-checking procedures increasingly available to users. This solution involves the provision of warnings that particular stories may be fabricated and takes advantage of crowdsourced fact-checking efforts; however, research has shown the limited impact of fact-checking efforts (Bernhard & Dohle, 2015). Finally, it is useful to detect the origin of fake stories and to filter out information from those sources (Ma et al., 2016). This approach is gaining increasing relevance given that blog accounts and a handful of websites are known to be substantially responsible for the prevalence of fake news.

2.1.3. The origin of fake news

Information about the root of fake news has far-reaching applications in regulating or preventing future outbreaks (Lokhov, Mézard, Ohta, & Zdeborová, 2014). However, little has been done to identify the origins and diffusion patterns whereas the prediction of epidemic spreading and promotion of viral content in networks has attracted significant attention (Louni & Subbalakshmi, 2014). The process of searching for the root of misinformation on social media is particularly challenging for numerous reasons. First, it is difficult to track down the original source of online information because it is often posted and published by authors whose identity is hidden or unknown. Second, unlike traditional publication outlets, where the citation of sources is the norm, social media dialogues often neglect standard protocols for reporting who says what. Finally, as social media content is often transformed through the process of sharing and spreading, it is hard to determine which version of media content is closest to the original. For these reasons, the traditional network approach that often relies on a single snapshot observation and ignores the evolutionary process has faced significant challenges in searching for the origins of information.

2.1.4. Spread of fake news

The flow of information on social media is structured by several factors, including users' interpersonal networks, algorithms, and users' psychological motivations (Bakshy et al., 2015). However, research has shown that these determinants tend to limit the flow across cross-cutting positions and contribute to the "echo chambers," where information flows among like-minded users, and "filter bubbles," where online content is controlled by algorithms reflecting users' prior choices (Pariser, 2016; Stroud, 2010). In the 2016 U.S. presidential election, partisans were more likely to discuss the election with "their side" rather than the opposition (Pew Research Center, 2016a).

Due to ideological homophily in online networks, misinformation seems indisputably accurate to social media users while fake news stories spread quickly among like-minded groups (Spohr, 2017). During the 2016 presidential election, Republicans tended to be more likely to believe that pro-Trump fake news was authentic and that Democrats had been deceived by pro-Clinton fake news (Allcott & Gentzkow, 2017). Partisans formed false consensus through the constant exposure to fake news shared within the virtual social circles (Flynn, Nyhan, & Reifler, 2017).

2.2. Network approach

2.2.1. Previous approach

Social network analysis has experienced enormous academic growth due to expanded computing technology and the availability of large-scale digital trace data. The social network approach has provided a useful method to understand how information travels and how individuals place themselves at the intersections of paths where information flows (González-Bailón & Wang, 2016). In a visualized practice, networks are assumed to reflect natural forms of organization, and nodes and links represent the players and relationships among them. Under this assumption, many studies have illustrated how things and ideas are disseminated and how key players shape the structure of the network (Cappella, 2017; Goel et al., 2015). For example, in the context of political discussion, it is important to identify who influences the visibility of certain topics on social media and who spreads such conversations within and beyond the network (Goel et al., 2015; Jürgens, Jungherr, & Schoen, 2011). Prior network analyses have also examined the influential source nodes for contagious viral messages (Dong, Zhang, & Tan, 2013).

Although this line of research has utilized an actor-based network approach, it has not fully benefited from a content-based network approach, which focuses on the evolutionary nature and hierarchy among media content. As actor-based analyses mostly provide a snapshot of an actor-based network, we continue to have limited understanding of how content evolves over the spreading process. It is a challenging procedure to incorporate the temporal dynamic of media content into the network, but it is particularly important for topics such as the spreading of rumors or epidemics (Jiang, Wen, Yu, Xiang, & Zhou, 2017).

2.2.2. Evolution tree analysis

Coming from the biological metaphor, an evolution tree, also known as a phylogenetic, is a tree graph that illustrates the evolutionary history and relationships among entities (Goldberg, Goldberg, Phillips, & Sorkin, 1996; Hsieh et al., 2015). Visually, an evolution tree shows a cluster of nodes and branches like other network graphs following the notion of graph theory. This visualization is especially effective when a network evolves from a starting point or needs to demonstrate a hierarchy. Except for the leaves on the bottom level, each node on the evolution tree is the most common ancestor of its descendants, and the edge between ancestor nodes and descendant nodes represents evolutionary distance, which may mean the evolution time of the mutation or number of the inserted nucleobase. The tree structure and the weight of the edges indicate the dissimilarity or distance among the entities in the evolutionary process. This information could assist researchers in understanding the evolutionary trends, and even predicting the further evolution. Previous research (Hsieh et al., 2015) showed that the evolutionary tree has the ability to expose similarities among entities during the evolutionary progress; such information could help researchers understand the evolutionary trends and even make future predictions. This insight inspires us that, in the context of fake news, we can build an evolution tree using media content about news stories (either fake or real) that will provide critical information about root content, content hierarchy, and content evolution.

The evolution tree approach has obtained a significant status for solving problems in various disciplines, such as sequence alignment (Parmentier, Trystram, & Zola, 2006) and the originality of digital files (Dias, Goldenstein, & Rocha, 2013). The principal function of an evolution tree in the digital era is to discover the modification and transformation relationship between a set of similar digital objects.

The evolution tree has been used in textual contexts to build a text phylogeny that can indicate the transformation history within a set of similar documents. Given that digital information is vulnerable to modification, reproduction, and dissemination, it is important to identify the original content and how this changes form over the spreading process. For example, if the text evolution tree of several near-duplicate documents can be successfully reconstructed, it could be useful in news tracking and plagiarism detection as well as finding intertextual connections between writers (Scheirer, Forstall, & Coffee, 2014). The researchers achieved reliable reconstruction of non-oriented trees especially for indirect edges of evolution tree with accuracy higher than 80%. Considering that our focus is on text-level content, the evolution tree suits our purposes of investigating the source and evolution of fake news tweets. The spreading pattern of a topic also helps understand the transformation and to identify the key event facilitating the transformation (Dias et al., 2013).

2.3. Present evolution analysis

This analysis aims to discover the evolutionary history of fake news' spreading on Twitter. The propagation of fake news often comes with a huge amount of retweeting and reposting after modification of the original message. With the assistance of Twitter properties such as content and timestamp, the evolution tree also helps identify the root of fake news.

The distance between a node and the root is the basis of the level. The number of levels in the tree is denoted as its depth, and the number of nodes in the level with the most nodes is denoted as the tree's breadth of the tree. In the tree of news spreading, whether fake or real, depth stands for how many times one topic has been modified or how many generations the topic has been derived. Breadth, on the other hand, represents the propagation scope of a topic. The degree of a vertex in a graph indicates the number of edges incident to the vertex. As the characteristics of a network can be described in terms of the depth, breadth, degree and their patterns, analyzing these features may help us understand the complicated network and distinguish between fake and real news stories.

On the foundation of the traditional evolution tree, we consider two critical factors in our evolution tree analysis: content-level similarity and timestamp information. The first key factor is calculated based on similarity among texts. The calculation of two input strings returns a number equivalent to the number of substitutions and deletions needed in order to transform one string into another. Second, unlike many other types of data such as modified similar images collected from the Internet, where there is no confirmed basis to indicate the file's date of creation, our Twitter dataset provides important metadata such as when the tweet was first posted. There is little chance that a tampered tweet could have an earlier timestamp than the original; thus, timestamp information is useful for tracking down the root of the evolution tree.

Therefore, by constructing and analyzing 60 evolution tree-based tweets collected on 30 fake and 30 real news topics, we ask the following questions:

RQ1. Who are the predominant root sources of fake news and real news?
RQ2. How do the evolution trees differ between fake and real news in terms of tree characteristics (i.e., breadth, depth, and degree)?
RQ3. How do evolution trees differ between fake news favoring Donald Trump and Hillary Clinton in terms of tree characteristics (i.e., breadth, depth, and degree)?

3. Method

3.1. Data source

Twitter data used in this study were obtained from the Crimson Hexagon database, which is a private firm providing social media content such as tweets, Facebook posts, and blog entries. Twitter's open access policy via application programming interface (API) has become increasingly restricted, which makes working with a firm such as Crimson Hexagon necessary to capture the entire archive (firehouse) of Twitter data. Crimson Hexagon provides a keyword-based search within specific user-defined periods, thereby allowing us to effectively gather tweets that meet our requirements. To examine fake and real news disseminated on Twitter, we selected a total of 60 news stories: 30 fake news and 30 real news stories. We chose news items that were widely circulated on the Internet before and after the 2016 US presidential election. The time period of the news stories was ranged from January 1, 2016, to April 30, 2017. To present fake news stories in a politically balanced way, we included 15 fake stories intended to damage Trump and another 15 fake stories intended to hurt Clinton.

We used the following criteria to include news items that were popular around the election period. First, to identify popular fake news, we explored a variety of sources, including newspapers, online news, and political blogs. For example, we investigated fake news stories reported by *Milwaukee Patch* (Anderson, 2016) and the *Mercury News* (May 2016). Some news articles (e.g., Silverman, 2016) provided a ranking of popular news stories shared on social media platforms during the election campaign. For example, we used widely-circulated fake news stories such as Pope Francis endorsing Donald Trump and Hillary Clinton selling weapons to ISIS. Additionally, we included real news stories that generated the most Facebook engagement throughout the election campaign (Silverman, 2016). Once we collected 100 news topics based on these rankings and media coverage, we double-checked the popularity of these news topics using Crimson Hexagon's database. We finalized our 60 news items based on the volume of tweets retrieved from the database. All news stories used in our study produced at least 500 tweets in the database. The actual number of tweets for each news item is indicated in Tables 1 and 2. For news stories generating more than 10,000 tweets, we randomly included 9999 tweets for our analysis.[1] To make sure we classified real and fake news accurately, we used two fact-checking websites (factcheck.org. And snopes.com). All fact-checking results confirmed our intentions.

3.2. String distance

In this paper, our approach focuses on text phylogeny and follows a similar method in biological phylogeny to create an evolution tree and express the evolutionary process of a set of tweet texts on the same topic. In order to build the evolution tree, we need to first calculate the string distance on every pair of tweets and combine them into a matrix. The algorithm that generates the evolution tree takes the matrix as input, and it outputs a minimum spanning tree that indicates all the string distances. We compared several widely employed methods for calculating string distance to decide the most suitable one for our purpose. After carefully testing options, including Levenshtein distance, Hamming distance, TF-IDF, and Q-gram, we chose Q-grams as our string distance algorithm.[2]

3.3. Analysis method

Our approach first computed the pairwise distance for the input data to construct the distance matrix and applied our algorithm to the distance matrix in order to generate the phylogenetic tree. The algorithm consisted of two steps:

1) We sorted all data into a priority queue in ascending order based on their date information. Therefore, the first (head) item in the priority queue should be the earliest data (out of all data items currently in the queue). We also initialized an empty tree.
2) We went through the priority queue iteratively, popping out the current head item from the priority queue in each iteration. The first data item was set as the root of the tree. For every subsequent data item, we examined all distances between current data and the data represented by each node in the tree and then grafted the current data to the node that had the minimum distance to the current data. Each iteration decreased the size of the priority queue by one, and the algorithm was terminated when the priority queue became empty.

When the phylogenetic tree of the whole dataset was fully constructed, we ran an algorithm to parse the tree and analyze its depth (i.e., the length of the longest path from the root of the tree down to a terminal node which has no child), breadth (i.e., the largest number of nodes whose path to the root node have the same length), maximum degree (i.e., the number of child nodes sharing the same parent node with the largest number of child nodes), and total degree (i.e., the sum of the numbers of every node's child nodes).[3]

4. Results

RQ1 examines the root content of each fake news tree and who posted the root tweet. Tables 1 and 2 provides information about 60 selected fake and real news stories. Table 1 shows that 87% of fake news tweets (26 out of 30) were first shared by ordinary users who were not known as journalists, celebrities, or politicians. Among the 30 fake news stories, only two were first tweeted by non-average users, and the other two accounts that first tweeted the news had already been suspended by Twitter.

In terms of the two news stories tweeted by non-ordinary users, one was with regard to voting machine problems in Utah while the other was with respect to Betsy DeVos's school gun policy speech. The former was first tweeted by President Donald Trump, who was the Republican presidential candidate at that time. On November 18, 2016, Trump mistakenly cited CNN saying that there were voting machine problems across the entire country, when Utah officials had made the official report. CNN later clarified that the problems had only been reported by one county in Utah, not the entire country as Trump had tweeted. However, Trump's tweet had gone viral before CNN could correct it.

The latter was first posted by David Siegel, a media coordinator at CNN, with a news link to the full story by CNN. DeVos, who is now Secretary of Education, stated “I would imagine there is probably a gun in a school in Wyoming to protect children from potential grizzlies” when answering school gun policy questions at her Senate confirmation hearing. CNN reported the story with the title, “Citing grizzlies, education nominee says states should determine school gun policies”; however, DeVos cited Mike Enzi, the Senator from Wyoming, instead of grizzlies.

Although most fake news stories were first tweeted by ordinary users, this does not necessarily indicate that those users first created fake news stories. It is more likely that those users first imported fake news information from somewhere else and spread it via Twitter. In line with this conjecture, 43% of fake news stories (13 of 30) were posted on Twitter with a link to non-credible news websites, such as ABCnews.com.co, a fake news website that

[1] We used this cut-off limit because some news items produced the enormous number of tweets. To make 60 news items reasonably comparable, we only included 9999 tweets for these stories.

[2] Levenshtein distance and Hamming distance are metrics for directly measuring the difference between two sequences. However, these two methods only count the number of substitutions, which may return inaccurate results for those pairs that share the same blocks at different positions (e.g., “TheHammingDistance” and “TheHamingDistance”). TF-IDF is a numerical statistic that is designed to model the importance of each term. When Cosine Similarity is incorporated, TF-IDF can serve as a measure of string similarity. However, the statistic of TF-IDF is derived from a corpus containing every term found in the dataset, which means if new data is added, the corpus must be updated and all TF-IDF values recalculated. This will be a burdensome task considering the huge size of vocabulary in the corpus. The Q-grams (Ukkonen, 1992) method would “slide” a window of length q to create a set of “grams” and the distance comes from two string sets instead of the string themselves. This feature makes Q-grams distance-insensitive to typographical error and format variation such as abbreviation. As for efficiency, Q-grams distance takes less time to calculate than Levenshtein distance, and does not require burdensome updates as does the TF-IDF method. Therefore, we chose Q-grams as our string distance algorithm.

[3] The detailed information about actual codes used is available upon request from the corresponding author.

Table 1
The characteristics of fake news trees.

Fake News (N)	Ideology	Contents	First source	Types	Posted Date	Followers	Following	News link attachment
1 (3060)	Pro Trump	folks supporting Islam Muslims if female should want a 25 to 50 Muslim raping as support for ISLAM ISIS Hillary 1st on TV	@thoughts54	Citizen/unaffiliated	1/9/16 0:40	217	483	No
2 (1493)	Pro Trump	Hillary wants to ban handguns & shut down the NRA if she becomes president, she'll protect us like she protected the Benghazi heroes!	@don_price60	Citizen/unaffiliated	1/9/16 3:18	2112	2050	No
3 (3917)	Pro Trump	Obama Admin: We Refuse To Accept North Korea As A Nuclear Power, Even Though It Is: Since taking office back i … https://t.co/aIrlStuOuY	@davegalant	Citizen/unaffiliated	1/11/16 19:58	41	99	Truth and Action, a news and opinion website with users loaded words frequently and often is not factual.
4 (6349)	Pro Trump	RT @WDFx2EU7 Hillary Campaign: Obama Is A Muslim, Wants To Ban Handguns, Does 'A Little Blow' #Wikileaks https://t.co/fw1jzDavQ0	Account suspended		1/18/16 1:55			No
5 (8239)	Pro Trump	#Palin endorses #Trump. Rather like the Pope endorsing Catholicism. Will the real 2016 candidate #pleasestandup?	@LynneDenee	Citizen/unaffiliated	1/19/16 21:17	794	1595	wtoe5news.com (not available)
6 (7423)	Pro Trump	Obama now is banning using the name of God in our pledge of allegiance. This man is a dictator. and America elected him twice very scary	@TomOrr777	Citizen/unaffiliated	2/14/16 5:35	5643	4982	No
7 (580)	Pro Clinton	After we declare independence we might look into becoming a Province of Canada. Will Trump build a wall around us?	@bsbafflesbrains	Citizen/unaffiliated	2/21/16 2:33	584	175	No
8 (1325)	Pro Trump	Tulsi Gabbard has to quit the DNC to support Bernie, but Cook Co Dems can mock up a fake ballot supporting Hillary?https://t.co/9Izqs6SM84	@the_roncarey	Citizen/unaffiliated	3/9/16 21:47	57	248	Daily Kos is a group blog and internet forum focused on liberal American politics.
9 (503)	Pro Clinton	YES! - RAGE AGAINST THE MACHINE To Reunite and Release Anti Donald Trump Album https://t.co/7ZWqtlaGFT via @Heavier Metal	@alyssascheinson	Citizen/unaffiliated	3/12/16 2:11	1299	467	Heavier Metal, a fake news site.
10 (9999)	Pro Clinton	RT @sheens_korner @realDonaldTrump Donald Trump Protester Speaks Out: "I Was Paid $3500 To Protest Trump's Rally" https://t.co/nKk65XW4E8 #Trump #Trump2016	@RJB_trading2	Citizen/unaffiliated	3/24/16 20:34	6156	6743	ABCnews.com.co, a fake news site which mimics the URL, design and logo of the actual news site ABCnews.com
11 (9356)	Pro Trump	@FBI Why Clinton Foundation received millions of dollars from Middle East before/after selling arms deal as SoS. https://t.co/VVa3WVoWLG	@LauraLRussell13	Citizen/unaffiliated	4/10/16 17:54	29	42	Foxnews
12 (4283)	Pro Trump	How severe could security concerns be with Hillary's email when the Pentagon still has nuclear launch codes on 8 inch floppy?	@NeoRenfield	Citizen/unaffiliated	5/26/16 14:47	5750	5287	No
13 (504)	Pro Clinton	Pope Francis Shocks World, Endorses Hillary Clinton for President, Releases Statement	@777Francejacque	Citizen/unaffiliated	7/10/16 0:44	3281	3812	wtoe5news.com (not available)
14 (9290)	Pro Trump	Hillary says she doesnt know anything about the leaked dnc emails, her private emails,benghazi, isis etc. y elect a potus that cant even email	@Spone63	Citizen/unaffiliated	7/25/16 21:36	1171	2097	No
15 (2908)	Pro Trump	RT @PatVPeters Khizr Khan's Deep Legal, Financial Connections to Saudi Arabia, Hillary's Clinton Foundation Tie Terror, Immigration https://t.co/1rol1QqyhO	@TarrAdam	Citizen/unaffiliated	8/1/16 21:42	380	646	Breitbart News Network is a far-right American news, opinion and commentary website.
16 (9999)	Pro Trump	@HuffingtonPost #wikileaks Hillary gave weapons to ISIS but MSM block out lol	@fresco_media	Citizen/unaffiliated	8/3/16 14:22	185	646	No
17 (4299)	Pro Clinton	RIGGED - Electronic Voting Machine hacked by Computer Scientists - YouTube https://t.co/hg5XewBZSs	No exists anymore	Citizen/unaffiliated	8/9/16 1:35			Youtube (the link is not available)

(continued on next page)

Table 1 (*continued*)

Fake News (N)	Ideology	Contents	First source	Types	Posted Date	Followers	Following	News link attachment
18 (1284)	Pro Clinton	Pence probably fist pumps every time he reads that a bullied trans youth committed suicide. Fuck Mike Pence. and fuck Trump. #Youngstown	@MHArcadia	Citizen/unaffiliated	8/15/16 19:37	190	44	No
19 (558)	Pro Clinton	Rupaul claims Trump touched him inappropriately in the 1990s - https://t.co/rCyJSLBEWW via @Shareaholic	@gloriajay	Citizen/unaffiliated	10/15/16 8:53	48	312	www.worldnewsdailyreport.com, a fake-news website or information and entertainment purposes only.
20 (3702)	Pro Clinton	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/F3bq0ldwQj WTF?	@mrclassicalmusi	Citizen/unaffiliated	10/19/16 16:55	468	1302	Newslo, located at Politicops.com, says it is the "first hybrid news/satire platform on the web.
21 (4135)	Pro Clinton	From White House to Interracial Bathhouse: Adventures of the Bacha Bazi Boys. Bust of terrorist MLK looks on as Western Civilisation burns.	@KalishJantzen	Citizen/unaffiliated	11/1/16 14:01	4265	4444	No
22 (7646)	Pro Clinton	BREAKING: ABC Uncovers MILLIONS Of Payments From Russia To Trump, Campaign Panics	@darlastraka	Citizen/unaffiliated	11/1/16 1:34	158	316	No
23 (6364)	Pro Trump	FBI Agent Suspected in Hillary Email Leaks Found Dead in Apparent Murder-Suicide https://t.co/b2nOzOaTVF	@w0z75	Citizen/unaffiliated	11/5/16 18:24	381	831	The Denver Guardian, a fake news website.
24 (9145)	Pro Trump	Just out according to @CNN: "Utah officials report voting machine problems across entire country"	@realDonaldTrump	POTUS	11/8/16 21:58	13089957	41	No
25 (788)	Pro Clinton	African Billionaire Will Give $1 Million To Anyone Who Wants To Leave America if Donald Trump is Elected President https://empireherald.com/back-to-africa-movement-to-help-blacks-leave-america-if-donald-trump-becomes-president/ …	@Babarfaridi__	Citizen/unaffiliated	11/9/16 6:25	132	52	Empire Herald, a fake news site.
26 (9145)	Pro Clinton	Pakistani TV: "Trump was born in Pakistan and not in America"! https://t.co/Qm69C9aWoz	@PamelaJaneVP	Citizen/unaffiliated	11/14/16 13:39	11641	12310	Jihad Watch, a website promotes an Islamophobic worldview and conspiracy theories.
27 (2152)	Pro Trump	@MDHillRaiser Trump Won 3084 of 3141 Counties, Clinton Won 57. Republic of Democratic states. 57 counties doesn't get to elect the President.	@Bill_Crowder_	Citizen/unaffiliated	11/15/16 20:58	7	18	No
28 (2228)	Pro Clinton	Y'all need to focus on Trump's cabinet not Kanye West …	@FEEovaErrythng	Citizen/unaffiliated	11/18/16 13:31	1369	1008	No
29 (1112)	Pro Clinton	#Trump Treasury pick Steven Mnuchin foreclosed on a 90-yr-old woman after a 27-cent payment error.	@RVAwonk	Citizen/unaffiliated	12/1/16 15:55	8034	3118	Politico.com
30 (972)	Pro Clinton	Citing grizzlies, DeVos tells Senator Murphy, rep of Sandy Hook, that states should determine school gun policies.	@dcsiegel	Media Coordinator at @CNN	1/18/17 2:02	623	513	CNN.com

mimics the logo and URL of ABCnews.com.

In contrast, 13 real news stories (43%) were initially triggered by journalists, news media, or political groups while 17 real news stories were first shared by ordinary users. It is not so surprising that many real news stories came from more credible sources such as journalists and news media outlets. More specifically, eight stories were first shared by journalists, including a senior editor of the *Huffington Post*, a White House editor of the *Washington Post*, and an anchor of FoxNews. Four stories were initially tweeted by news organization outlets such as FoxNews, San Francisco News Now, Reuters Politics, and Reuters Top News. As a political group, New Hampshire Tea Party first produced a real news story.

These 13 real news stories included certain news links. Among them, 10 real news stories delivered news links of mainstream media such as nytimes.com, washingtonpost.com and FoxNews.com and online news media such as huffingtonpost.com. Two stories displayed links to professional blogs, including lawfareblog.com and thegatewaypundit.com. Interestingly, one news story involved a screenshot of print newspapers. In addition, these news links were shared by five ordinary users, five journalists, and three news media outlets. News media and their journalists tended to tweet real news stories by using more credible news websites.

Moreover, we checked whether any source of tweets in our analysis had been generated by bots or cyborg accounts. BotOrNot (Davis, Varol, Ferrara, Flammini, & Menczer, 2016) provides a useful framework for determining whether a tweet was posted by human or bot. When a username is fed into Python API (https://github.com/truthy/botornot-python), the framework analyzes the user's profile, spanning content and other metadata, and returns a score that suggests the likelihood that the suspected account is indeed a bot. A strong bond was not found between bot accounts and fake news sources, indicating that most of the sources are humans

Table 2
The characteristics of real news trees.

Real News (N)	Contents	First source	Types	Posted Date	Followers	Following	News link attachment
1 (9351)	@KatiePavlich @Mediaite: Biden can run in 2020! or run as VP in 2016!	@hanschoimd	Citizen/unaffiliated	1/7/16 04:46	37	214	No
2 (843)	"RT @XplodingUnicorn [Shooter is Muslim] Trump: Ban Muslims [Shooter is an immigrant] Trump: Ban immigrants [Shooter is white] Trump: Ban Muslim immigrants"	@TheGamerShaggy	Citizen/unaffiliated	1/10/16 12:55	598	1454	No
3 (7015)	RT @HatefMokhtar Nearly 100 killed in Russian air strike in Syria's Idlib https://t.co/BTdwdFuo7g	@iaslesixmusicla	Citizen/unaffiliated	1/11/16 04:28	82	1175	Theoslotimes.com
4 (1298)	So the ego maniac Trump bullies Scottish to build golf course as usual he calls them losers for fighting him is this what we want 4 POTUS NO	@patsman1949	Citizen/unaffiliated	1/19/16 11:31	404	901	No
5 (9999)	#Breaking: A federal judge in #Seattle says the state of #Washington does not have standing to sue over #Trump immigration ban (via @Reuters)	@KevRincon	Anchor/Reporter at FoxNews	2/3/17 23:42	1061	295	No
6 (1369)	Trump declines to attend White House Correspondent's Dinner. Draft Alec Baldwin for Trump's Understudy!	@rkvukovic	Citizen/unaffiliated	2/25/17 22:54	113	146	No
7 (2261)	Emails Warrant No New Action Against Clinton, Comey Says https://t.co/6iizhUgDen via @nytimes	@sdmattpotter	Senior editor at the SD Reader	2/25/17	120 K	117 K	Nytimes.com
8 (9999)	@Rabidtwinkie77 i gave you multiple facts. Obama los Independents but won election in 2008. more Americans voted for Hillary than Bernie	@patrick102977 (Now @BigPat1029)	Citizen/unaffiliated	3/17/16 22:47	1358	4976	No
9 (1356)	I missed the SGA debate but sounds like one of the candidates just did the old duck and dodge tactic. Been watching trump in the debates?	@LincolnO1995	Citizen/unaffiliated	3/23/16 01:37	377	398	No
10 (1282)	Karns softball will be hosting Clinton Thursday at 6:00 for an Alzheimer's awareness game! Come out and support and wear your purple!	@Myia_beeler	Citizen/unaffiliated	4/16/16 4:06	308	303	No
11 (9658)	Ann Coulter Appearance At UC Berkeley Canceled Over Security Concerns http://dlvr.it/Nwy8gL #sanfrancisco	@SFnewsnow	San Francisco News Now	4/19/17 19:00	12.3 K	410	Sanfrancisco.cbslocal.com
12 (9999)	RT @BuzzFeedNews Voters Honored Susan B. Anthony By Placing Their "I voted" Stickers On Her Grave In New York https://t.co/kpXppkkUxB https://t.co/jTUHXLxjfT	@risenc	Deputy editor, NYT op-ed page	4/20/16 20:41	4364	2668	Buzzfeed.com
13 (2021)	MORE: Trump wants to renegotiate or terminate 'horrible' trade deal with South Korea, wants it to pay for $1-billion THAAD missile defense	@ReutersPolitics	Reuters Politics	4/28/17 01:42	239 K	394	No
14 (1832)	Top officials @ Democratic National Committee privately planned how to undermine Sanders's according to a trove of emails released by WikiLe	@WeTheNorth11	Citizen/unaffiliated	7/22/16 21:36	3681	1971	No
15 (9999)	@Susan_Hennessey: Russian efforts to influence our democratic processes is the real story. https://t.co/IY42Y5nKRC https://t.co/Vb71WLXtxh	@LSKSDave	Citizen/unaffiliated	7/25/16 12:10	283	287	Lawfareblog.com (Blog)
16 (8629)	RT @thehill New York Post publishes nude photo of Melania Trump on cover https://t.co/IjFBjApMyd https://t.co/7nvrzdbpel	@rickeybolin	Citizen/unaffiliated	8/1/16 00:00	2830	4043	The Hill news
17 (9176)	No real American patriot would support either Hillary or trump they are both CIA controlled the CIA is the illuminati's police force	@NHTeaParties	New Hampshire Tea Parties	8/2/16 02:15	3834	2254	No
18 (4027)	lol. I hate double standards. Prosecute Hillary and undo an injust! YOU have to Stop pretending! https://t.co/80qvPUrHfc	@becky_avon3985	Citizen/unaffiliated	8/4/16 7:05	855	189	No
19 (9373)	@WPalisoul FAR WORSE THAN TRUMP ≫ History of corruption': Clinton Cash documentary author Peter Schweizer https://t.co/EMP0G3YGlw	@mrtom3560	Citizen/unaffiliated	8/7/16 1:38	3111	4791	RT News (YouTube)

(continued on next page)

Table 2 (*continued*)

Real News (N)	Contents	First source	Types	Posted Date	Followers	Following	News link attachment
20 (2764)	Kellyanne Conway new role is technically @realDonaldTrump \Campaign Manager\ but her true title is, "The Trump Whisperer." Watch it unfold	@TheBrodyFile	Chief Political Correspondent for CBN News	8/17/16 03:53	10.7 K	1079	No
21 (9999)	Trump recorded having extremely lewd conversation about women in 2005 https://t.co/Te5hNQYBtL	@DanEggenWPost	WP White House editor	10/7/16 13:02	11.9 K	356	Washingtonpost.com
22 (9999)	FBI is looking at Anthony Weiner's laptop. Let's hope they're wearing hazmat suits.	@marcwilmore	Citizen/unaffiliated	10/28/16 21:33	2124	325	No
23 (2795)	.@SenRonJohnson: \President Obamaâ€⟨ was emailing Secretary Clinton at ClintonEmail dot com. He had to know.\ "#SundayFutures @MariaBartiromo https://t.co/uSe1O61WAa"	@FoxNews	FoxNews	10/30/16 07:24	15.1 M	440	FoxNews.com
24 (9999)	@realDonaldTrump Trump was GREAT at MCC College yesterday in Michigan! The line went all the way thru campus and out! TRUMP WINS MICHIGAN!	@terrybaas1	Citizen/unaffiliated	11/1/16 06:41	30	23	No
25 (7060)	.@Beyonce: "We have the opportunity to create more change. I want my daughter to grow up seeing a woman lead our country".	@danmericaCNN	Political producer @CNN	11/5/16 02:02	36.4 M	1120	No
26 (4097)	"I wouldn't give this dolt the time of day, but ….Kanye West Meets TRUMP at Trump Tower https://t.co/2U4q4qXYz6"	@SteveC_USA (Now @PatriotSteve_US)	Citizen/unaffiliated	12/13/16 14:42	3464	3254	Thegatewaypundit.com (political blog)
27 (7701)	Bill Clinton on Trump \He doesn't know much. One thing he does know is how to get angry white men to vote for him, \ "https://t.co/j544cMcJqv"	@MddeH	Geopolitics reporter @qz	12/19/16 11:48	972	2841	Screen shot of print newspaper
28 (2910)	Final popular vote total shows Clinton won almost 3 million more ballots than Trump https://t.co/OmLtjH7Rr1	@nickpwing	Senior Viral Editor @HuffPost	12/20/16 22:33	6565	1255	Huffingtonpost.com
29 (2019)	Trump: Sprint to bring back 5000 jobs to US	@codyest92	Citizen/unaffiliated	12/28/16 22:09	317	287	No
30 (1151)	Trump on North Korea ICBM: 'It won't happen!' https://t.co/39a80hnN3i	@Reuters	Reuters Top News	1/2/17 23:39	18.2 M	1054	Reuters.com

according to the likelihood acquired from the API.

RQ2 investigates the differences between fake news and real news trees in terms of evolution tree characteristics—namely, breadth, depth, and degree. The breadth and depth represent the width and height of the tree, respectively. Based on the evolution tree adjusted by content-level similarity and timestamping, the way we get the values of the breadth and depth is by traversing the entire tree and recording the number of nodes in the level having the largest number of nodes as the breadth as well as by recording the length of the longest path from the root to any terminal node as depth. The relevance of breadth is that a tree with larger breadth indicates that the non-altered news content spreads out more

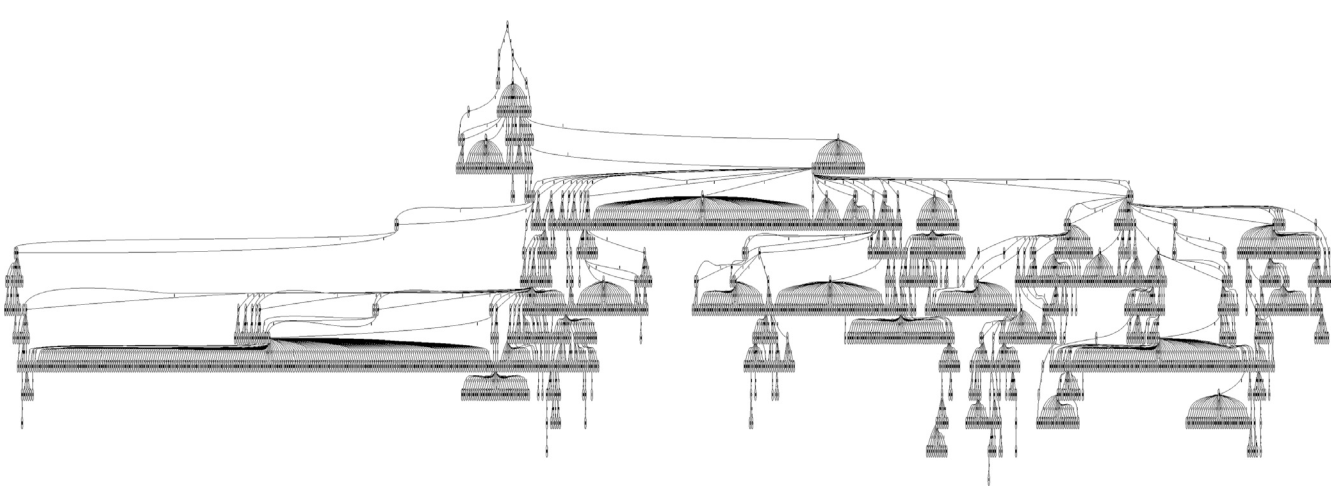

Fig. 1. Evolution tree of tweets about fake news #14 (Trump was born in Pakistan).

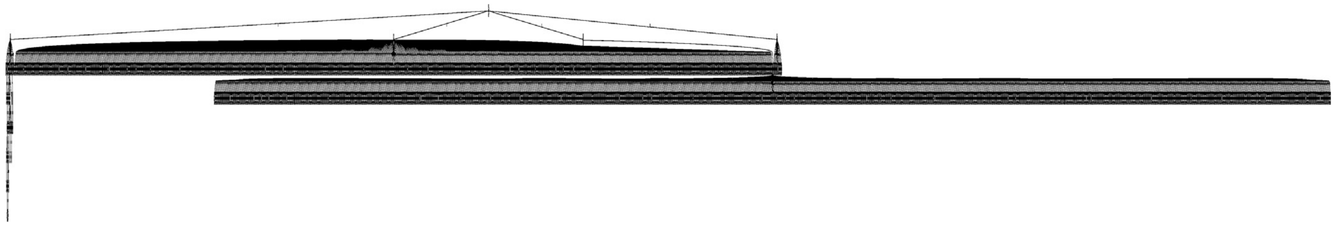

Fig. 2. Evolution tree of tweets about real news #17 (Obama was emailing Secretary Clinton at clintonemail.com).

widely. In the case of a tree with larger breadth, it is more likely that other users just retweet exactly the same content. On the other hand, the relevance of depth is that a tree with larger depth indicates that news content is subjected to more alterations during its spread.

The *t*-test results show that the average breadth of fake news ($M = 5.95$, $SD = 1.79$) is smaller than that of real news ($M = 7.83$, $SD = 4.12$, $t = -2.48$, $p < .05$). On the other hand, the average depth of fake news ($M = 2.07$, $SD = 0.76$, $t = 2.46$, $p < .05$) is greater than that of real news ($M = 1.67$, $SD = 0.56$). These differences in the characteristic of evolution trees are visually illustrated in Fig. 1 (fake news #14) and Fig. 2 (real news #17). For example, the evolution tree in Fig. 1 appears to display greater vertical depth, suggesting that the fake story has undergone a myriad of modifications by many users over the spreading process. This also means that the modification of content did not occur uniformly at the same time. On the other hand, the evolution tree in Fig. 2 displays wider breadth but smaller vertical depth, suggesting that the news story spread to Twitter users without any modification of news content and that this spreading process happened almost simultaneously. This makes sense given that real news items are typically broadcast rapidly and widely by a great number of news outlets, and social media users share the stories without adding much to the original content. On the other hand, when sharing fake news stories, social media users tend to modify the original content at their discretion. This may be due in part to users having some suspicion of the source's reliability or the content itself.

As the number of news stories on each topic are not exactly the same, we think the normalized degree of each topic may be more persuasive than the total degrees. The average normalized degree of real news ($k = 30.55$) and fake news ($k = 20.92$) is not statistically different from our sample size ($k = 20.92$, $t = 0.75$, $p = .46$). All 60 evolution tree analyses are available upon request from the corresponding author.

RQ3 asks how the trees differ between fake news favoring Donald Trump and fake news favoring Hillary Clinton. According to *t*-test results, neither the average breadth nor depth of fake news favoring Trump ($M = 6.20$, $SD = 1.85$; $M = 2.11$, $SD = 0.64$; $t = 0.81$, $p = .44$) are significantly different from those favoring Clinton ($M = 5.70$, $SD = 1.45$; $M = 2.04$, $SD = 0.88$, $t = 0.22$, $p = .83$). Similarly, the normalized degree of those topics favoring Trump ($k = 22.12$) and Clinton ($k = 19.73$, $t = 0.14$, $p = .89$) have no considerable disparity.

5. Discussion

The recent rise in fake news during the 2016 US presidential election warrants the development of a mechanism to regulate the flow of fake news on social media. One approach involves identifying the source of fake news and controlling information from those sources. This study addressed this issue by paying attention to the fact that content changes form as it spreads. Using the methodological framework of evolution tree analysis, this study searched for the antecedent content of fake news tweets and evaluated the evolution patterns of fake and real news stories. This study used a computational network science approach to ask questions about important and impending issues that the fields of communication and journalism are facing.

Findings from 60 evolution tree analyses showed that most tweets about fake news were generated by accounts from ordinary users. This is somewhat divergent from recent findings that a handful of fake news websites are dominating the stream of misinformation (Shao et al., 2016). Given that it should be increasingly difficult to monitor and control a panoply of ordinary accounts instead of a few suspects, the finding may pose additional challenges for efforts to design regulatory algorithms. However, it should be noted that half of the tweets with misinformation included a link to non-credible news websites. Thus, it may not be that fake news websites directly promote fake stories via Twitter, but rather that ordinary users bring fake news from those websites to the Twittersphere. These findings suggest that we should continue to keep an eye on non-credible news websites as well as the link between those websites and social media users who share fake news from them.

The reliance on non-credible sources online supports the notion that there has been a significant change in the patterns of individuals' news consumption and the traditional role of the journalists as gatekeepers (Coddington & Holton, 2014; Jang & Park, 2017). Audiences' heavy use of digital media may weaken the function of traditional media in the dissemination and verification of information. With the explosive growth of media outlets, the public is now able to bypass traditional gatekeepers and gain direct access to online information sources (Westerman, Spence, & Van Der Heide, 2014). Without the protection of gatekeepers and the ability to detect misinformation, the public becomes susceptible to fake news, particularly news items generated by websites that use deceptive strategies to resemble authentic news outlets (Weissman, 2016).

Our study found significant differences in the network characteristics between real and fake news content. On average, real news trees showed greater breadth whereas fake news trees displayed greater depth. In the evolution tree analysis, greater depth indicates a greater number of modifications in content, and greater breadth indicates a wider dispersion. One interpretation of these findings is that tweets about real news stories are shared and retweeted without many revisions, but tweets about fake stories undergo more modification. When users spread factual news stories, it seems that users share or retweet factual news without commenting on the content or modifying the information. Additionally, as real news stories are typically tweeted immediately after the news stories are published online in news websites, tweets about them are shared and retweeted in a very short time. These sharing patterns may lead to greater breadth and shorter depth for tweeting real news stories. On the other hand, fake news stories appear to be circulated more slowly than factual news on social media, and users are more likely to attach their own opinions when sharing. This may imply that at least some Twitter users tend to process fake news stories with greater caution, and eventually end up with non-sharing or significant revision before sharing. It should also be noted that fake news stories that are not based on real-world events do not hold much relevance to one of most critical elements of real news stories: timeliness. Therefore, fake news stories may have a longer life expectancy, and their buzz can be revived at any time.

One interesting example in our analysis was a fake news tweet from President Trump. The evolution tree showed the widest breadth but the shortest depth among 60 news trees, indicating that Twitter users simply adopted and forwarded President Trump's tweet without modifying it. Although scholars disagree about the impacts opinion leaders have on the diffusion of information (e.g., Stieglitz & Dang-Xuan, 2012; Zhang, Zhao, & Xu, 2016), our study highlights the role of politicians in creating popular politics-related trends in the online environment.

It is important to address several shortcomings of this study. First, although we first demonstrated that the framework of the evolution tree network could offer valuable insights into the origins and changing aspects of misinformation in the Twittersphere, it was challenging to evaluate the framework's performance accurately. Second, our analysis focused on the difference between fake and real news at an aggregated level, but little attention was paid to the linguistic perspectives. It would be interesting to study how language evolves over the sharing process. Third, our analysis involved 60 news items, which hardly covered a wide range of news topics. Future research should expand the focus further by examining other misinformation contexts, such as the link between vaccines and autism or climate change hoax frames, as well as by analyzing a larger scale dataset. Other types of datasets can also be investigated, such as news articles captured from news websites. Fourth, as there is no limitation on the length of the articles, the method of calculating the distance may vary. We also intend to utilize some metadata information from tweets to assist with the reconstruction of the evolution tree and make it closer to the actual evolution. For example, the retweet function in Twitter was used to decide the closest ancestor if more than one node featured the same content. Finally, our findings showed important differences between tweets about fake and real news stories, but they are far from sufficient for inducing a general regulatory mechanism. Additional efforts to identify the key characteristics of fake news are needed to improve the quality of information flow in digital space.

6. Conclusion

In this project, our team of social scientists and computer science engineering researchers proposed a novel framework for analyzing, identifying, and predicting the flow of misinformation on digital platforms. Using an evolution tree modeling method, we examined how misinformation was initiated, transmitted, and managed and how it evolved in the hybrid online news system. Specifically, we identified antecedent tweets about fake news stories as well as the authors of those root tweets. We also observed that tweets about real and fake news showed different evolution patterns.

Moving forward, to combat fake news, academics and industries need richer insights into the presence and spread of misinformation on social media venues. In this vein, the current findings help broaden our understanding of the origin and spreading pattern of fake news stories, which is essential to enhance the quality of information flow in digital space. Given that prior research has raised doubts about the role of fact-checking efforts and crowdsourcing corrections (Flynn et al., 2017), it is important to understand socio-technical factors embedded in a networked system and to design interventions and algorithms that disrupt the flow of information from non-credible sources. For example, by detecting information promoted by non-credible sources or related bot accounts, computer scientists can create algorithms that decrease the visibility of such information. This approach would be particularly effective if a handful of sources are linked to the origin of most fake information. Finally, along with this line of algorithmic approach, educational efforts, including literacy interventions about information and digital media, can help inform the digital public about the issues at hand and minimize the misperceptions surrounding them (Mele et al., 2017).

References

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives, 2*(31), 211–236. https://doi.org/10.1257/jep.31.2.211.

Anderson, S. (2016). *Top 20 'Fake News' stories of the 2016 presidential election according to Buzzfeed analysis*. The Milwaukee Patch. Retrieved from https://patch.com/wisconsin/milwaukee/top-20-fake-news-stories-2016-presidential-election-according-buzzfeed-analysis.

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science, 348*(6239), 1130–1132. https://doi.org/10.1126/science.aaa1160.

Bernhard, U., & Dohle, M. (2015). Corrective or confirmative actions? Political online participation as a consequence of presumed media influences in election campaigns. *Journal of Information Technology & Politics, 12*(3), 285–302. https://doi.org/10.1080/19331681.2015.1048918.

Cappella, J. N. (2017). Vectors into the future of mass and interpersonal communication research: Big data, social media, and computational social science. *Human Communication Research, 43*, 545–558. https://doi.org/10.1111/hcre.12114.

Coddington, M., & Holton, A. E. (2014). When the gates swing open: Examining network gatekeeping in a social media setting. *Mass Communication & Society, 17*(2), 236–257. https://doi.org/10.1080/15205436.2013.779717.

Davis, C. A., Varol, O., Ferrara, E., Flammini, A., & Menczer, F. (2016). Botornot: A system to evaluate social bots. In *Proceedings of the 25th international conference companion on world wide web* (pp. 273–274). International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/2872518.2889302.

Dias, Z., Goldenstein, S., & Rocha, A. (2013). Exploring heuristic and optimum branching algorithms for image phylogeny. *Journal of Visual Communication and Image Representation, 24*(7), 1124–1134. https://doi.org/10.1016/j.jvcir.2013.07.011.

Dong, W., Zhang, W., & Tan, C. W. (2013). Rooting out the rumor culprit from suspects. In *Paper presented at the Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*. https://doi.org/10.1109/ISIT.2013.6620711.

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM, 59*(7), 96–104. https://doi.org/10.1145/2818717.

Flynn, D. J., Nyhan, B., & Reifler, J. (2017). The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Political Psychology, 38*(S1), 127–150. https://doi.org/10.1111/pops.12394.

Goel, S., Anderson, A., Hofman, J., & Watts, D. J. (2015). The structural virtuality of online diffusion. *Management Science, 62*(1), 180–196. https://doi.org/10.1287/mnsc.2015.2158.

Goldberg, L. A., Goldberg, P. W., Phillips, C. A., & Sorkin, G. B. (1996). Constructing computer virus phylogenies. In *Annual symposium on combinatorial pattern matching* (pp. 253–270). Berlin, Heidelberg: Springer. https://doi.org/10.1007/3-540-61258-0_19.

González-Bailón, S., & Wang, N. (2016). Networked discontent: The anatomy of protest campaigns in social media. *Social Networks, 44*, 95–104. https://doi.org/10.1016/j.socnet.2015.07.003.

Hsieh, S., Tsai, I., Hung, H., Chen, Y., Chou, H., & Lee, C. (2015). An enhanced algorithm for reconstructing a phylogenetic tree based on the tree rearrangement and maximum likelihood method. *Intelligent Computing Theories and Methodologies Lecture Notes in Computer Science*, 530–541. https://doi.org/10.1007/978-3-319-22186-1_53.

Jang, S. M., & Kim, J. K. (2018). Third-person effects of fake news: Fake news regulation and media literacy interventions. *Computers in Human Behavior, 80*, 295–302.

Jang, S. M., & Park, Y. J. (2017). Redirecting the focus of the agenda: Testing the zero-sum dynamics of media attention in news and user-generated media. *International Journal of Communication*, 3998–4017.

Jiang, J., Wen, S., Yu, S., Xiang, Y., & Zhou, W. (2017). Identifying propagation sources in networks: State-of-the-art and comparative studies. *IEEE Communications Surveys & Tutorials, 19*(1), 465–481. https://doi.org/10.1109/COMST.2016.2615098.

Jin, Z., Cao, J., Guo, H., Zhang, Y., Wang, Y., & Luo, J. (2017). Detection and analysis of 2016 US presidential election related rumors on Twitter. In *International*

conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation (pp. 14–24). Cham: Springer.
Jürgens, P., Jungherr, A., & Schoen, H. (2011). Small worlds with a difference: New gatekeepers and the filtering of political information on Twitter. In *Proceedings of the 3rd international web science conference* (p. 21). ACM. https://doi.org/10.1145/2527031.2527034.
Lokhov, A. Y., Mézard, M., Ohta, H., & Zdeborová, L. (2014). Inferring the origin of an epidemic with a dynamic message-passing algorithm. *Physical Review E, 90*(1), 012801. https://doi.org/10.1103/PhysRevE.90.012801.
Louni, A., & Subbalakshmi, K. P. (2014). A two-stage algorithm to estimate the source of information diffusion in social media networks. In *Computer communications workshops (INFOCOM WKSHPS), 2014 IEEE conference on* (pp. 329–333). IEEE. https://doi.org/10.1109/INFCOMW.2014.6849253.
Ma, J., Gao, W., Mitra, P., Kwon, S., Jansen, B. J., Wong, K. F., et al. (2016). *Detecting rumors from microblogs with recurrent neural networks* (pp. 3818–3824). In IJCAI.
May, P. (2016). *Did some of these fake news stories help send Trump to the White House? The Mercury News*. http://www.mercurynews.com/2016/11/11/did-fake-news-on-facebook-send-trump-to-the-white-house/.
Mele, N., Lazer, D., Baum, M., Grinberg, N., Friedland, L., Joseph, K., et al. (2017). *Combating fake news: An agenda for research and action*. https://shorensteincenter.org/wp-content/uploads/2017/05/Combating-Fake-NewsAgenda-for-Research-1.pdf.
Mihailidis, P., & Viotty, S. (2017). Spreadable spectacle in digital culture: Civic expression, fake news, and the role of media literacies in "post-fact" society. *American Behavioral Scientist, 4*(61), 441–454. https://doi.org/o0r.g1/107.171/0770/0200726746241271770011217.
Pariser, E. (2016). *The filter bubble: How the new personalized web is changing what we read and how we think*. New York, NY: Penguin.
Parmentier, G., Trystram, D., & Zola, J. (2006). Large scale multiple sequence alignment with simultaneous phylogeny inference. *Journal of Parallel and Distributed Computing, 66*(12), 1534–1545. https://doi.org/10.1016/j.jpdc.2006.03.003.
Pennycook, G., Cannon, T. D., & Rand, D. G. (2017). *Implausibility and illusory truth: Prior exposure increases perceived accuracy of fake news but has no effect on entirely implausible statements*. Available at: SSRN https://ssrn.com/abstract=2958246.
Pew Research Center. (2016a). *How Americans are talking about Trump's election in 6 charts*. Retrieved from http://www.pewresearch.org/fact-tank/2016/12/22/how-americans-are-talking-about-trumps-election-in-6-charts/.
Pew Research Center. (2016b). *Many Americans believe fake news is sowing confusion*. Retrieved from http://www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/.
Scheirer, W., Forstall, C., & Coffee, N. (2014). The sense of a connection: Automatic tracing of intertextuality by meaning. *Literary and Linguistic Computing, 31*(1), 204–217. https://doi.org/10.1093/llc/fqu058.
Shao, C., Ciampaglia, G. L., Flammini, A., & Menczer, F. (2016). Hoaxy: A platform for tracking online misinformation. In *Proceedings of the 25th international conference companion on world wide web* (pp. 745–750). International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/2872518.2890098.
Silverman, C. (2016). *This analysis shows how viral fake election news stories outperformed real news on Facebook*. Buzzfeed News. Retrieved from https://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook?utm_term=.jj13ePW3Eb%23.wnZeJaye47.
Spohr, D. (2017). Fake news and ideological polarization: Filter bubbles and selective exposure on social media. *Business Information Review, 34*(3), 150–160. https://doi.org/10.1177/0266382117722446.
Stieglitz, S., & Dang-Xuan, L. (2012). Political communication and influence through microblogging—An empirical analysis of sentiment in Twitter messages and retweet behavior. In *System science (HICSS), 2012 45th Hawaii international conference on* (pp. 3500–3509). IEEE. https://doi.org/10.1109/HICSS.2012.476.
Stroud, N. J. (2010). Polarization and partisan selective exposure. *Journal of Communication, 60*(3), 556–576. https://doi.org/10.1111/j.1460-2466.2010.01497.x.
Sunstein, C. (2009). *On rumors: How falsehoods spread, why we believe them, what can be done*. New York: Farrar, Straus, Giroux.
Ukkonen, E. (1992). Approximate string-matching with q-grams and maximal matches. *Theoretical Computer Science, 92*, 191–211.
Weissman, C. G. (2016). *Watch out for this fake news website masquerading as the New York Times. Businessinsider.com*. http://www.businessinsider.com/nytimescomco-posts-fake-news-articles-pretending-to-be-the-new-york-times-2015-6.
Westerman, D., Spence, P. R., & Van Der Heide, B. (2014). Social media as information source: Recency of updates and credibility of information. *Journal of Computer-Mediated Communication, 19*(2), 171–183. https://doi.org/10.1111/jcc4.12041.
Zhang, L., Zhao, J., & Xu, K. (2016). Who creates trends in online social media: The crowd or opinion leaders? *Journal of Computer-Mediated Communication, 21*(1), 1–16. https://doi.org/10.1111/jcc4.12145.

Dr. S. Mo Jang (PhD, University of Michigan) is an assistant professor in the School of Journalism and Mass Communications, University of South Carolina. His research focuses on the flow of information in digital communication and public opinion about controversial political/science issues.

Tieming Geng is a Ph.D. student in the Department of Computer Science and Engineering, University of South Carolina. His research focuses on network security, network protocol and network transmission.

Jo-Yun Queenie Li is a Ph.D. candidate in the School of Journalism and Mass Communications at the University of South Carolina. Her primary research interest is health communication in diverse populations. One aspect of her research focus is about how mass communication can be used to promote beneficial changes in behavior among members of different populations.

Ruofan Xia is a Ph.D. candidate in the Department of Computer Science and Engineering, University of South Carolina. His research is focusing on phylogenetic analysis from genome rearrangement and cancer copy number variations.

Dr. Chin-Tser Huang is an associate professor in the Department of Computer Science and Engineering, University of South Carolina. His research focuses on network communication and network security. His research has been supported by NSF, DARPA, and AFRL.

Dr. Hwalbin Kim is a research professor in the Healthcare Media Research Institute at Hallym University, South Korea. His research interests include health communication, public opinion, and social media.

Dr. Jijun Tang is a professor in the Department of Computer Science and Engineering, University of South Carolina, and in the School of Computer Science and Technology, Tianjin University, China. His main research interests are algorithm, evolution and bioinformatics.