Searching for 'submarine' propaganda outlets: an efficient method to identify sources of false information on the web

Kohei Watanabe
Waseda University

Paper presented at EPSA Annual Conference 2019, Belfast

Computer scientists and technology companies are developing automated systems for to evaluate trustworthiness of pieces of information to prevent 'fake news' from spreading, but these computer systems' ability to perform fact checking independently of human supervision is so far very limited. In this paper, I prose an efficient method to identify sources of false information on the web using commercial search engines and an open-source software package for document similarity computation. Using Russia's state-controlled media as the known sources of false information, I not only found a 'satellite' website that actively disseminate articles published by Russian sources, but also 'submarine' websites that are professionally tailored to disguise themselves as independent news media based in European cities or the US states but remain low-profile by not even having official Twitter account. I argue that these submarine websites could be used as disposable propaganda outlets in future political events.

The frequent mentions of "misinformation" on the internet in recent publications by the Word Economic Forum (Howell, 2013) and the Organization for Economic Co-operation and Development (OECD) (Koulolias, Jonathan, Fernandez, & Sotirchos, 2018) seem to suggest that the spread of 'fake news' on the internet has been widely recognized as one of the most serious problems facing the developed countries. Although the term was used to refer to various types of documents ranging from news parodies to advertisements in early discussions, scholars agree that its essence is high facticity of the content with intention to deceive audiences (Rubin, Chen, & Conroy, 2015; Shu, Sliva, Wang, Tang, & Liu, 2017; Tandoc, Lim, & Ling, 2018; Volkova & Jang, 2018). The main concern expressed in these reports is the spread of false information by domestic or foreign actors for political purposes such as influencing elections results or policy making

processes. To counter such strategic misinformation, the European Union has allocated funds to develop a website and an automated system for fact checking.[1]

According to Duke Reporters' Lab that maintains a database of fact-checking projects, there are around 150 groups or organizations that publicly evaluate truthfulness of information (Stencel & Griffin, 2018). Although fact checking in these projects is conducted mostly manually, computer scientists and technology companies have been developing systems that automate fact checking to scale up the effort. However, computer systems' ability to perform such tasks independently of human supervision is so far very limited, because fact checking require software programs to perform series of complex tasks: extracting implicit claims from sentences, understanding social contexts of statements, synthesizing the evidence from different sources, and weighing claims based on their importance (Graves, 2018).

To avoid the technological complexity in evaluating individual pieces of information, Lazer et al. (2018) suggest detecting sources of false information instead. Detection of sources is not only easier for computer programmes but also more useful from policy perspective and consistent with the definition that 'fake news' is different from random mistakes: it allows policymakers to prevent spread of false information more easily by blocking the sources that repeatedly publish false information. In this paper, I demonstrate an efficient methodology to implement this approach at low cost targeting the Russian governments 'satellite' websites that is used to spread false information originating from Russian's state control media. This methodology is a combination of commercial search engine and an open-source software package, both of which are high accessible.

With the new methodology, I successfully detected News Front, a website listed as a pro-Russian fake news site by the European Union, as a satellite of RT, and discovered unknown websites that publish news articles supplied by RT in a very similar manner. My further investigation of the websites releveled that these websites often (1) hide real identity of their owners, (2) present themselves as local media outlets in Europe or the United States, and (3) remain inactive on social media lacking Twitter account. These characteristics suggest that they are 'fake news' websites that remain low-profile until significant political event that require active

[1] EU vs. Disinformation (https://euvsdisinfo.eu) and Co-Inform (https://coinform.eu).

propaganda. I propose to call such websites 'submarine' website as oppose to 'satellite' websites and urge political science scholars and fact-checking practitioners to monitor their contents.

Detecting spread of false information

Computer scientists have developed automated systems that combine natural language processing and machine learning techniques to detect the spread of false information on the internet (Graves, 2018; Thorne & Vlachos, 2018). Many of the systems are based on either (1) manually collected examples, (2) knowledge databases, (3) stylistic features, or (4) user behaviour to detecting pieces of false information. In the example-based approach, human coders assess trustworthiness of documents to train supervised machine learning models to replicate their judgement in larger scales (e.g. Castillo, Mendoza, & Poblete, 2011; Volkova & Jang, 2018). In the knowledge-based approach, the computer programmes extract truth claims from documents and asses their accuracy using databases of factual information (e.g. Dong et al., 2015; Julià et al., 2018). In the stylistic feature-based approach, stylistic features of documents are used to detect documents written to misleading audiences (e.g. Volkova & Jang, 2018). In the user behaviour-based approach, systems detect spread of false information based on social media users' reaction to certain messages (e.g. Del Vicario, Quattrociocchi, Scala, & Zollo, 2019; Wu & Liu, 2018).

However, only the example-based approach, which requires extensive human involvement to construct large dataset to train computer algorithms, has been successful (Graves, 2018). Even with the state-of-art natural language processing and machine learning techniques, it is difficult to extract factual claims from complex sentences, or stylistic features of false statements from elaborated documents (Graves, 2018; Volkova & Jang, 2018; Wu & Liu, 2018). The user behaviour-based systems are more robust as they do not rely on analysis of the content of documents, but we cannot prevent false information from spreading effectively in this approach.

Other group of computer scientists have developed techniques to detect source of false information based on (1) attributes of websites or (2) structures of hyperlinks on the web. Abbasi et al. (2010) attempted to detect 'fake website' based on texts, source code, URLs, images, hyperlinks in websites using a supervised machine learning model. Gyöngyi et al. (2004) and Woloszyn et al. (2018) developed algorithms similar to Google's Page Rank that discover spam websites or unreliable news websites based on hyperlinks. As semi-supervised algorithms, their systems only require a list of untrustworthy websites to estimate detect source of false information.

Methodology

Collecting webpage snippets

I subscribed to RSS feeds of three Russian state-controlled news websites (RT, TASS, and Sputnik News) and collected headings and summaries of their articles. I created a program that submit search queries created from the headings to the Bing Web Search API to retrieved web pages index within the last 24 hours.[2] For example, a heading "US Fails to Make Assad Trade Ties With Iran for Washington Support – Report" becomes a query string "us fails make assad trade ties iran washington support report". With such machine-generated search queries, I retrieved top 20 pages in each search and saved their URL, titles and descriptions, and the total number of hits. Between 7 November 2018 and 2 April 2018, the program executed the search every hour using 29,002 news summaries from the RSS feeds and collected 276,849 webpage snippets from 13,077 internet domains (Table 1).[3]

Table 1: Number of news summaries and webpage snippets

	TASS	RT	Sputnik	Total
News summary	6,131	7,247	15,624	29,002
Webpage snippet	48,768	86,380	141,701	276,849

Finding correlated outlets

I identified websites that publish news articles similar to the Russian outlets by computing similarity between the snippets and summaries. I formed bigrams from the texts to reduce ambiguity of features, and the sizes of resulting document-feature matrix became 276,849 × 1,355,940 for the snippets and 29,002 × 38,066 for the summaries. First, I computed cosine similarities between all the snippets and summaries; second, I counted the number of similarity scores larger than 0.1; third I weighted the counts by the inverse of the total number of hits before taking the sum for each internet domain.

---

[2] We started developing the system in July 2017 but discovered malfunctioning of Bing API's time search (called "freshness"). We reported the issue to Microsoft, but need to wait for 9 months until the problem was finally solved by its engineers in April 2018.

[3] The cost of the Bing Search API in this study was 100 to 200 GBP per month, depending on the number of search queries submitted.

Computing similarity between highly sparse and large DFMs usually requires expensive computational infrastructure but I achieved this analysis on a laptop computer by developing a highly efficient algorithm and implementing it as described below.

Algorithm

Cosine similarity has been widely used to detect text reuse in quantitative text analysis. Cosine similarity s between two vectors, x and y, who have the equal number of elements that record frequency of words, is define as:

$$cosine\ similarity = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}}$$

For a document-feature matrix, **X** and **Y**, with m and n columns, we can compute similarities of all the pairs of documents very quickly by matrix multiplication:

$$S_{kl} = (\mathbf{X} \times \mathbf{Y}) \circ \left(\sqrt{\mathrm{col}(X)^2} \times \sqrt{\mathrm{col}(Y)^2}^{\mathrm{T}}\right)^{-1}$$

where, S is a $m \times n$ similarity matrix, $\mathrm{col}(X)$ and $\mathrm{col}(Y)$ are column sums of **X** or **Y**, and $\circ$ denotes element-wise multiplication.

However, the size of S often becomes too large for computers that have small RAM when there are many documents. For example, the number of similarity scores in S will be 100,000,000, when both **X** and **Y** comprise of 10,000 documents. The number of documents depends on studies, but it can easily exceed 1 million if they are microblog posts or sentences of news articles.

My solution to the quadratic increase in the size of storage in RAM is flooring small similarity values, which are usually not of our interest, before constructing a full similarity matrix. In this algorithm, I break down S into an array of similarity vectors and apply a threshold t to floor smaller values to zero:

$$\acute{S}_{kl} = [\acute{s}_1, \acute{s}_2, \ldots, \acute{s}_m]$$

$$\acute{s}_k = \begin{cases} 0, & s_k < t \\ s_k, & s_k \geq t \end{cases}$$

It computes each similarity vector by multiplying a kth column vector $X_{\cdot k}$ from **X** by **Y**:

$$s_k = (X_{\cdot k}\mathbf{Y}) \circ \left(\sqrt{\mathrm{col}(X_{\cdot k})^2} \times \sqrt{\mathrm{col}(Y)^2}^{\mathrm{T}}\right)^{-1}$$

If the floored similarity matrix $\acute{S}$ is saved as a sparse matrix, which records only non-zero values, the required size of storage becomes smaller dramatically.

Implementation

I implemented the algorithm in a parallelized C++ function, where $\acute{s}_k$ is computed independently and concurrently, as part of the proxyC package using Intel's TBB Library. To demonstrate the advantage of the algorithm, I compared my function with a function in the proxy package by applying them to 99% sparse document-feature matrices (DFMs) generated random values for 1,000 or 10,000 documents on a laptop computer with a Core i7 processor. Figure 1 shows that proxyC is roughly 10 times faster than proxy with a DFM with 10,000 documents ("10k"), and 100 times faster with 1,000 documents ("1k") thanks to the low-level parallel computing.

Figure 1: Comparison between proxyC and proxy with DFMs in different sizes

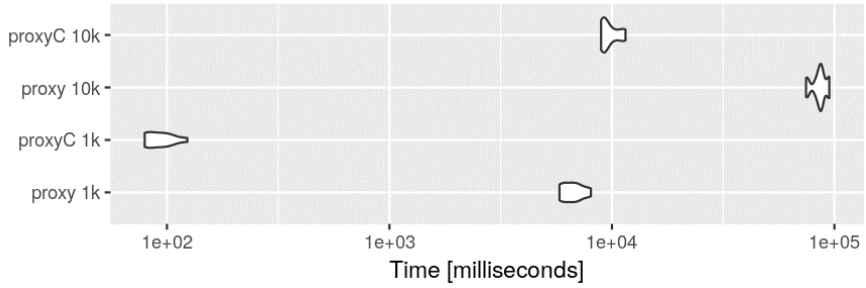
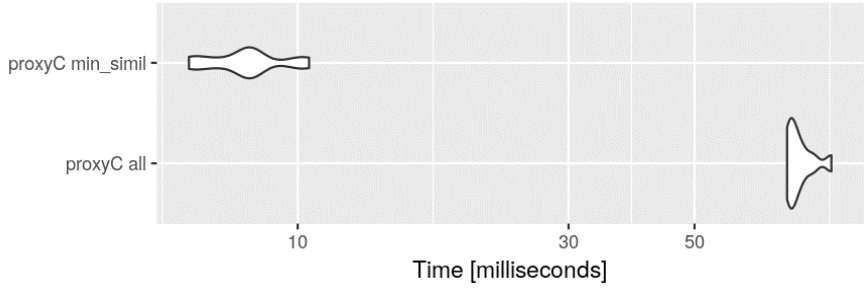


Figure 2: proxyC with or without minimum threshold



Further, a similarity matrix with all the scores for 10,000 documents is 762MB but it becomes 0.2MB when scores are floored with $t = 0.9$. Minimum threshold not only reduces the size of storage required, but also accelerates the function roughly 10 times, because it is faster create a similarity matrix from fewer values (Figure 2).

Clustering websites

I intended to downweight websites by the total number of hits because some of the news headings contain very common words, but I failed to exclude established websites (e.g. Wikipedia, CNN, Washington Post) that have many pages, which accidentally contain words used in headings in Russian news websites. To achieve this, I performed clustering of websites based on bigrams that occur both in the news summaries and webpage snippets in each search. For each internet domain, I grouped documents by search identifier, computed feature similarities of the two DFMs, and took diagonal elements greater than zero to find correlated bigrams. I used these bigrams as categorical variable to perform correspondence analysis. Clustering of websites by correlated features help us to separate satellite websites from others, because satellite websites correlates with Russian websites in specific topics.

Results

I initially performed website clustering with the summaries and snippets from all the Russian sources, but the result was most clear when only top-100 websites retrieved by search queries produced from RT's news headings. In Figure 3, mainstream news websites such as CNBC (www.cnbc.com), Washington Post (www.washingtonpost.com), Politico (www.politico.com) and CNN (www.cnn.com) appear low in the first dimension (left), while less known websites are high on this scale (right). Among those high on the first dimension but low in the second dimension (bottom-right), I find News Front (en.news-front.info), a pro-Russian website known as one of the major sources of false information in the EU's fact checking project.[4] Websites positioned high on the second dimensions are Current Affairs Online (currentaffairsonline.co.uk), International Africa (international-africa.com) Berlin Tomorrow (berlintomorrow.com) and Westminster Report (westminsterreport.com), none of which have been recognized as a source of false information in fact checking projects to author's knowledge.

[4] A blog post on the EU vs. Disinformation (https://euvsdisinfo.eu/no-news-on-the-news-front) explains that the website is run by pro-Russian activist in east Ukraine.

Figure 3: Correspondence analysis based on correlated bigrams

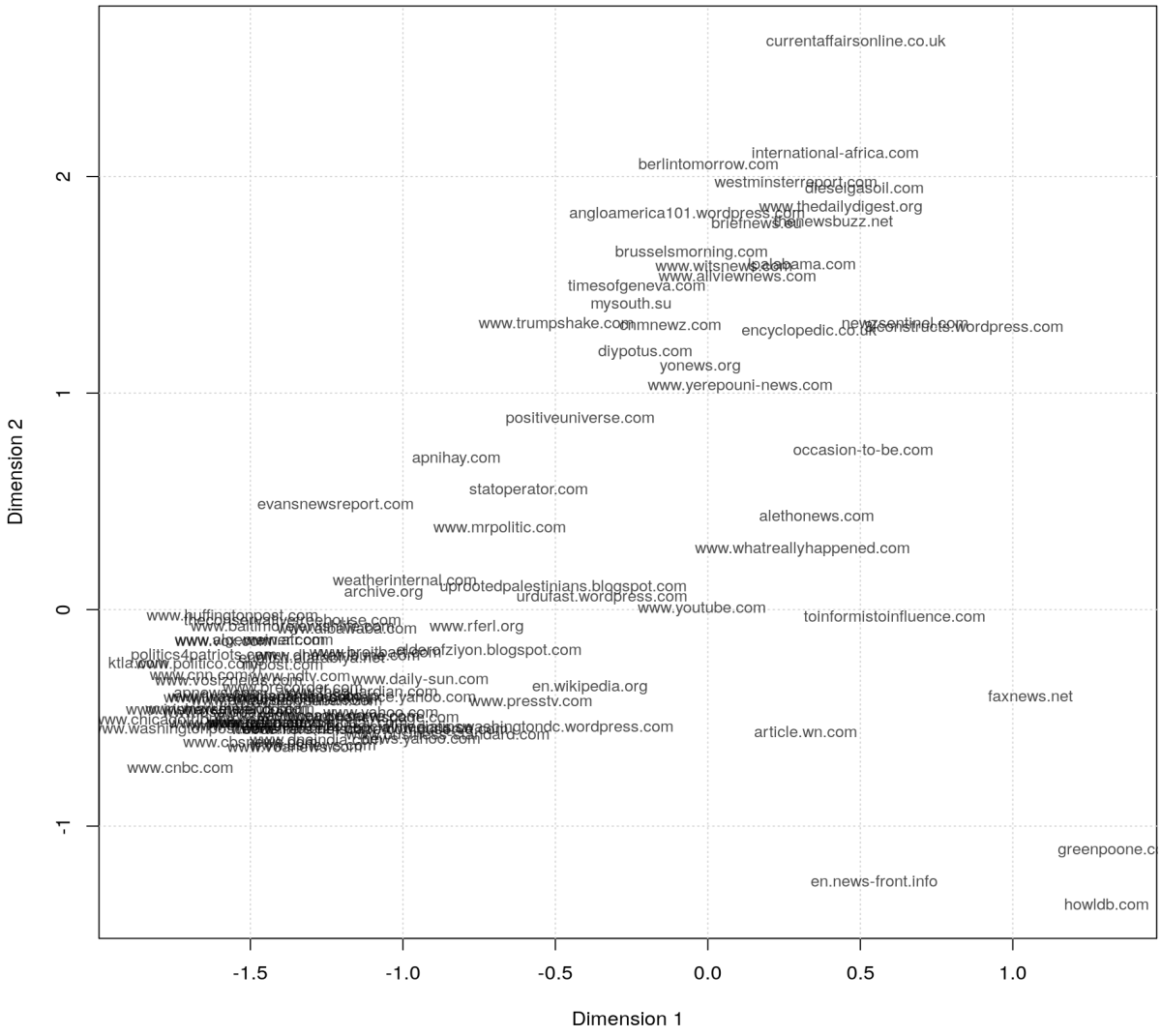


Table 2: Characteristics of top-20 websites on the second dimension

Name	Domain	Registrar	Registered Date	Author Identity	Twitter Account	Place Name	Ads
Current Affairs Online	currentaffairsonline.co.uk	godaddy.com	2017-09-25	-	Caffairsonline	-	3
International Africa	international-africa.com	godaddy.com	2017-09-25	-	-	Africa	0
Berlin Tomorrow	berlintomorrow.com	godaddy.com	2017-11-22	-	-	Berlin	1
Westminster Report	westminsterreport.com	godaddy.com	2017-11-22	-	-	Westminster	0
Diesel Gasoil	dieselgasoil.com	godaddy.com	2016-03-22	Nasneft Group (Russia)	-	-	4
The Daily Digest	www.thedailydigest.org	domain.com	2008-12-21	Offstream Media (US)	TheDailyDigest	-	3
Living in Anglo-America	angloamerica101.wordpress.com	-	-	-	Angloliving	-	0
The News Buzz	thenewsbuzz.net	namesilo.com	2016-09-28	-	-	-	2
European Brief News	briefnews.eu	wedos.com	2017-01-19	-	lilushkas	Europe	2
Brussels Morning	brusselsmorning.com	godaddy.com	2017-09-25	-	-	Brussels	2
Liberty Project Alabama	lpalabama.com	namecheap.com	2017-12-15	Magbook (US)	-	Alabama	0
Witsnews	www.witsnews.com	namecheap.com	2016-11-04	-	witsnews	-	0
All View News	www.allviewnews.com	epik.com	2017-11-04	-	-	-	1
Times of Geneva	timesofgeneva.com	godaddy.com	2017-08-01	Times of Geneva (Swizerland)	TimesofGeneva	Geneva	0
World News	mysouth.su	r01.ru	2010-05-14	-	-	-	1
Newz Sentinel	newzsentinel.com	namecheap.com	2015-10-15	-	-	-	2
Trump Shake	www.trumpshake.com	namecheap.com	2016-11-23	-	trumpshake	-	2
Christian News Menute	cnmnewz.com	fastdomain.com	2013-02-25	-	CNMNewz	-	2
3 i constructs	3iconstructs.wordpress.com	-	-	-	-	-	0
Encyclopedic News	encyclopedic.co.uk	lcn.com	2016-07-27	-	encyclopedicuk	-	4

Following the journalistic procedure to check authenticity of information on the internet (Sliverman, 2014), I studied characteristics of the top-20 websites on the second dimension qualitatively (Table 2). These websites have distinctive look, but all of them are built on the WorldPress platform; 18 websites have original domain names, many of which were registered to GoDaddy (godaddy.com) or Namecheap (namecheap.com) in late 2017; 9 websites have professionally designed logo marks (Figure 4). Only half of the website have Twitter account, even though it became very common to online news outlets to promote their news articles on social media in recent years. The number of advertisements on their front page was also between zero to four, suggesting that many of the outlets are not for commercial purposes. Six websites present themselves as if they are regional or local news media (Africa, Berlin, Westminster, Brussels, Alabama, Geneva), but there is no evidence that they have any connection to these places. Further, only four websites have information about website owners' identity. According to their self-identification, Diesel Gasoil is owned by a Russian energy company, Nasneft; The Daily Digest is affiliated with an independent US news website, Offstream Media; Liberty Project Alabama is New York-based publisher, Magbook; the Times of Geneva is an online edition of a Geneva-based print newspaper. However, it is likely that the identity of the owners of the last two websites are false, because the postal address of Magbook is the same as the company that sells WordPress templates, and there is no newspaper called the *Times of Geneva* in Switzerland.[5]

Discussion

I found the unknow websites are disguising as regional or local news outlets with professionally designed logo marks and hiding the real identity of their owners. They are "concocted" websites that presenting themselves as a legitimate outlet by combining several elements (Abbasi et al., 2010). According to Fogg (2003), main factors that determine credibility of websites are (1) design look, (2) information structure, (3) information focus, and (4) author's motives. Websites built with the WordPress platform and professional logo marks can easily meet

[5] The Time of Geneva website claims that "Approaching 35 years in business, Times of Geneva is a well-known and respected name in Geneva, Switzerland, with an increasingly recognizable media brand across Switzerland and around the world".

the first two criteria. All those unknown websites mix stories from RT and other sources to meet the last two criteria too.

Although names of the unknown websites contain reference to places, the most obvious false identity is found in the Times of Geneva, which present itself as an online edition of newspaper established in the city. However, claiming links to well-known cities is an effective strategy of false news websites because internet users usually find it difficult to verify identity of organizations behind websites. The false identification as a newspaper based in Geneva, the home to the United Nations agencies, is aimed at enhancing credibility of its international news stories. It is not easy for me to reveal the groups or individuals behind false news websites, but the website is likely linked to China, because many of the stories are supplied by its satellite news broadcaster, CGTN, representing the Communist Party's views on international affairs.

Very little effort that International Africa, Berlin Tomorrow, Westminster Report, and Brussels Morning is making to earn advertisement revenue or to promote themselves on social media indicates their unique function in spreading false information on the internet. My initial goal was to discover 'satellite' websites that actively spread information originate from Russia's state-controlled media, but such active involvement in international propaganda attract attentions of fact checkers who will blacklist such websites. Therefore, it is better for propaganda websites to remain low-profile until significant political events occur, and active misinformation becomes necessary. When they involve in propaganda, social media would be fully utilized to reach broader audiences. Creation of websites prior to occurrences of such political events is not only to respond quickly but also to enhance credibility of the websites. This type of propaganda outlets can be called 'submarine' websites as oppose to 'satellite' websites.

Although there is no direct evidence that suggest links between the unknown websites and the Russian government, there is a reason for propagandist states to employ submarine websites. When credibility of news outlets is one of the most important factors that determines the effectiveness of propaganda, propagandist states must invest much resources to establish credible images of state-controlled outlets, but extensive use of such outlets in propaganda campaigns erodes their credibility (Snegovaya, 2015). Therefore, a better strategy for them is building disposable submarine website at low costs and spreading false information on social media using submarine websites as independent source that support their claims. As soon as social media users

or fact checkers notice that these are 'fake news' websites, the propagandist can abandon them and deploy other submarine websites to continue its campaign.

Further, my analysis of the websites by reverse DNS resolution revealed that Europe Brief News, Current Affairs Online, Brussels Morning and Berlin Tomorrow are hosted on the same web server along with other 16 news websites (Table 3). These websites refer to countries (Germany, France, Britain, America, Australia), the US states (Washington DC, California, Florida), and European cities (London, Cambridge, Brussels, Berlin, Amsterdam) in their names and falsely identify themselves as independent news media based in these places. This suggest that reference to place is a common strategy for false news websites to enhance their credibility.

Figure 4: Logo marks of unknown websites

Table 3: False news websites hosted on the same web server

Name	Domain	Place names	Top-20
Germany Latest News	germanylatest.com	Germany	
France News 7	francenews7.com	France	
Fox News Online 24	foxnews24.info	-	
Florida Report Daily	floridareportdaily.com	Florida	
Extra London News	extralondon.co.uk	London	
Extra American News	extraamerican.com	America	
Evening Washington News	eveningwashington.com	Washington	
Europe Brief News	europebriefnews.com	Europe	✓
Edition Online	editiononline.co.uk	-	
Edition America	editionamerica.com	America	
The Daily Cambridge	dailycambridge.co.uk	Cambridge	
Current Affairs Online	currentaffairsonline.co.uk	-	✓
California Today	californiatoday.net	California	
Brussels Morning	brusselsmorning.com	Brussels	✓
Britain Today News	britaintodaynews.com	Britain	
Britain Post News	britainpost.co.uk	Britain	
Business News Report	bnreport.com	-	
Berlin Tomorrow	berlintomorrow.com	Berlin	✓
Australia News Today	australianews.today	Australia	
Amsterdam Times	amsterdamtimes.info	Amsterdam	

Conclusions

The use of search engine alone to collect webpage snippets that resemble to articles published by known source of false information (e.g. RT, TASS, and Sputnik) appeared ineffective because of large number of false hits, but we can refine the result by compute document similarities and temporal correlation between the webpage snippets and news summaries. Although computing similarities between large number of documents requires a large storage space in RAM that is much larger than laptop computers usually have, I achieved this by developing an efficient software program: it reduces the size of storage space in RAM to fractions of original sizes by applying minimum threads hold. Thanks to the smaller output object and parallel computing, it also accelerates similarity calculation by 10 to 100 times. I made this program publicly available as an R package, proxyC, on the Comprehensive R Archive Network (CRAN).

With the new methodology that combines search engine and the new software program, I successfully identified News Front as a RT's satellite website. More interesting, I also found previously unknown websites with professional design look and reference to well-known

European cities but without information about website owners and official social media accounts. I propose to call these outlets 'submarine' websites as oppose to 'satellite' website, because they remain low-profile until significant political events that require active propaganda. Considering their common characteristics and their potential utility as a propaganda outlet, I argue that is organized effort to install submarine websites by groups or individuals that affiliated with propagandist states such as Russia and China.

My success in discovering the submarine websites by the simple procedure suggest that development of computer systems that discover *sources* of false information is more effective than those that detect *pieces* of false information. When fact-checking systems that automatically evaluate truthfulness of individual claims have only limited capacity, it is more efficient to assess trustworthiness of the origins of the claims. If a list of suspicious websites is created, fact checkers can monitor articles published in the websites or linked from social media posts, or even block them to prevent false information from spreading on the internet.

I focused on false information that originate from Russia's state-controlled media, but the methodology presented in this paper should not be restricted to this. Since this methodology only requires a website a source of information, we can discover, for example, websites run by the Saudi or Chinese government or religious or racial extremist groups. Since there are too many problematic sources of information on the web today, I hope other political science scholars and fact-checking practitioners to embark of similar studies using the tool that I made publicly available as an R package.

References

Abbasi, A., Zhang, Z., Zimbra, D., Chen, H., & Nunamaker, J. F. (2010). Detecting Fake Websites: The Contribution of Statistical Learning Theory. *MIS Quarterly*, *34*(3), 435–461. https://doi.org/10.2307/25750686

Castillo, C., Mendoza, M., & Poblete, B. (2011). Information Credibility on Twitter. *Proceedings of the 20th International Conference on World Wide Web*, 675–684. https://doi.org/10.1145/1963405.1963500

Del Vicario, M., Quattrociocchi, W., Scala, A., & Zollo, F. (2019). Polarization and Fake News: Early Warning of Potential Misinformation Targets. *ACM Trans. Web*, *13*(2), 10:1–10:22. https://doi.org/10.1145/3316809

Dong, X. L., Gabrilovich, E., Murphy, K., Dang, V., Horn, W., Lugaresi, C., … Zhang, W. (2015). Knowledge-based Trust: Estimating the Trustworthiness of Web Sources. *Proceedings of the VLDB Endowment*, *8*(9), 938–949. https://doi.org/10.14778/2777598.2777603

Fogg, B. J., Soohoo, C., Danielson, D. R., Marable, L., Stanford, J., & Tauber, E. R. (2003). How Do Users Evaluate the Credibility of Web Sites?: A Study with over 2,500 Participants. *Proceedings of the 2003 Conference on Designing for User Experiences*, 1–15. https://doi.org/10.1145/997078.997097

Graves, L. (2018). *Understanding the promise and limits of automated fact-checking*. Reuters Institute.

Gyöngyi, Z., Garcia-Molina, H., & Pedersen, J. (2004). Combating Web Spam with Trustrank. *Proceedings of the Thirtieth International Conference on Very Large Data Bases - Volume 30*, 576–587. Retrieved from http://dl.acm.org/citation.cfm?id=1316689.1316740

Howell, L. (2013). *Global Risks 2013*. World Economic Forum.

Julià, P., Denaux, R., Farrell, T., Sjöberg, C. M., Nenzén, S., Rodríguez-Pérez, A., … Shah, S. I. (2018). *Generic Co-Inform architecture* (No. Version 2). Retrieved from Co-Inform website: https://coinform.eu/wp-content/uploads/2019/01/D4.2-H2020-Co-Inform-Generic-Architecture.pdf

Koulolias, V., Jonathan, G. M., Fernandez, M., & Sotirchos, D. (2018). *Combating Misinformation: An Ecosystem in Co-creation*. Retrieved from The International Council for Information Technology in Government Administration website: http://ica-it.org/images/publications/Combating-misinformation.pdf

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., … Zittrain, J. L. (2018). The science of fake news. *Science*, *359*(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

Rubin, V. L., Chen, Y., & Conroy, N. J. (2015). Deception detection for news: three types of fakes. *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community*, 83. American Society for Information Science.

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, *19*(1), 22–36.

Sliverman, C. (2014). *Verification handbook*. European Journalism Centre.

Snegovaya, M. (2015). *Putin's information warfare in Ukraine* (No. 1). Retrieved from The Institute for the Study of War website: http://www.understandingwar.org/sites/default/files/Russian%20Report%201%20Putin%27s%20Information%20Warfare%20in%20Ukraine-%20Soviet%20Origins%20of%20Russias%20Hybrid%20Warfare.pdf

Stencel, M., & Griffin, R. (2018, August 7). The number of fact-checkers around the world: 156... and growing. Retrieved May 8, 2019, from Duke Reporters' Lab website: https://reporterslab.org/the-number-of-fact-checkers-around-the-world-156-and-growing/

Tandoc, E. C. J., Lim, Z. W., & Ling, R. (2018). Defining "Fake News." *Digital Journalism*, *6*(2), 137–153. https://doi.org/10.1080/21670811.2017.1360143

Thorne, J., & Vlachos, A. (2018). Automated fact checking: Task formulations, methods and future directions. *ArXiv Preprint ArXiv:1806.07687*.

Volkova, S., & Jang, J. Y. (2018). Misleading or Falsification: Inferring Deceptive Strategies and Types in Online News and Social Media. *Companion Proceedings of the The Web Conference 2018*, 575–583. https://doi.org/10.1145/3184558.3188728

Woloszyn, V., & Nejdl, W. (2018). DistrustRank: Spotting False News Domains. *Proceedings of the 10th ACM Conference on Web Science*, 221–228. https://doi.org/10.1145/3201064.3201083

Wu, L., & Liu, H. (2018). Tracing fake-news footprints: Characterizing social media messages by how they propagate. *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, 637–645. ACM.