EPiC Series in Computing

Volume 93, 2023, Pages 141–151

Proceedings of Society 5.0 Conference 2023

Uncovering Cross-Platform Spreading Patterns of Fake News about Covid-19

Lukas Schiesser[1] and Hans Friedrich Witschel[1] and Andre De La Harpe[2]
[1] FHNW University of Applied Sciences and Arts Northwestern Switzerland
[2] Cape Peninsula University of Technology

Abstract

The spreading of fake news or misinformation on social media is a serious threat to modern societies, making more and more people susceptible to being unfairly influenced in their decision-making, be it in elections or other democratic processes. We contribute to the body of work in the area of fake news detection by studying *cross-platform,* multivariate spreading patterns of fake news on Covid-19-related topics – where existing studies have focused strongly on single platforms and/or on single metrics or indicators. Our findings show that there are several attributes that are specific to the cross-platform spreading process that become important predictors of fake news: there is e.g. a clear tendency that fake news travels faster from one platform to the other than real news. Meanwhile, although we have compiled a cross-platform corpus of fake and real news that future research may build on, data availability remains a challenge for future work.

1 Introduction

It is broadly agreed that malicious spreading of misinformation poses a threat to modern societies and democracies, e.g. when voters get influenced by misleading information in election campaigns (Zhang & Ghorbani, 2020). In democracies, misinformation regarding the "political elites" can be harmful beyond the behavior in single elections, in the sense that they can create a general apathy or mistrust in politics, or encourage extremism (Lazer et al., 2018).

It has been shown that "fake news" is re-posted more often, by more people and more rapidly than true information (Lazer et al., 2018). In terms of propagation networks, Zhou & Zafarani (2019) have found that fake news have deeper propagation trees and that propagation networks are denser than true news networks – where these effects are magnified by the use of social bots (Lazer et al., 2018). These observations underline the magnitude of the threat that fake news is posing.

Here, the promise of Society 5.0 – solving social challenges with the help of (technological) innovations – comes into play: as discussed e.g. by Haron (2022) in the context of the Malaysian Society 5.0, advanced information technologies should be capable of empowering citizens to better judge the trustworthiness of information and thus solve a pressing issue of modern society.

Consequently, the detection of misinformation and fake news has received great attention in the data mining community. Various approaches are being discussed, e.g. based on identifying misinformation-specific language patterns (Oshikawa et al., 2018), visual presentation and images (Parikh & Atrey, 2018), as well as context factors such as evaluation of author/spreader credibility and of network-related spreading patterns (Zhou & Zafarani, 2020).

Given the relatively large success of context-based approaches, based on e.g. investigating user-comment graphs (Liao et al., 2021) or social engagement of users (Shu, Wang, & Liu, 2017), it is surprising that most existing studies focus on spreading of fake news *within* a single social media platform. In fact, because of the high accessibility of data, a large number of studies focuses on Twitter, often even using the same dataset, e.g. (Liao et al., 2021; Ren & Zhang, 2021; Sitaula et al., 2020). What has been studied far less is the spreading of fake news *across* a variety of platforms. Several authors express their belief that such cross-platform analyses could shed a new light on how fake news spreads and evolves (Wilson & Starbird, 2020; Yang et al., 2021; Zhou & Zafarani, 2020). However, recent studies attempting to study cross-platform spreading of fake news (Chen et al., 2021; Papakyriakopoulos et al., 2020) are a) hitherto constrained to measuring certain pre-defined metrics, mostly related to the virality of spreading and b) pointing out that “Data access remains a bottleneck” (Chen et al., 2021). The data access problem refers to the challenge of reliably identifying the same news story on different platforms, but also to the collection of contextual attributes, such as identifying authors on those platforms – a research field on its own (Shu, Wang, Tang, et al., 2017).

In this study, our goal is to shed some light onto the spreading patterns of fake news across various social media platforms. Specifically, we investigate *multivariate* patterns of spreading processes that occur when (mis-)information spreads across platforms, i.e. we go beyond studying isolated metrics to describe that process. In doing so, we do not determine the relevance of attributes/metrics in advance but try to use a broad range of them (limited of course by the data access problems mentioned above), such that the degree of relevance of the various attributes for detecting fake news across platforms also emerges as a result of the data mining process.

We contribute to the common body of knowledge on two levels: one level presents the identified cross-platform patterns where we a) make further advances in overcoming constraints of measuring virality-related pre-defined metrics (Chen et al., 2021; Papakyriakopoulos et al., 2020) and b) go beyond such simple metrics by providing *multivariate spreading patterns*. The other level tackles the data foundation for further research as data availability is a challenge in this field of research. By providing a data set that identifies the same information across platforms we support further research.

2 Related Work

Misinformation is difficult to detect using solely content-based approaches as the news is often created to mislead recipients (Shu, Silva, Wang, et al., 2017). Using style analysis for example only could not solve the issue of fake news (Potthast et al., 2017). The criticality of the time factor favors context-based misinformation detection approaches as disproof of past news might fail to support disproving current news (Ren & Zhang, 2021). Which is why fighting misinformation and fostering true and qualitatively valuable content, requires the understanding of the origin and spreading process of information (Chen et al., 2021). Source-based detection approaches focus on evaluating the credibility of authors and spreaders, which in the social media context, is mostly identifying bots or human users likely to spread misinformation (Zhou & Zafarani, 2020). Using a hybrid approach of content detection and source-based detection past associations with misinformation and the number of different authors an article shows could be proven a significant factor in detecting fake news (Sitaula et al., 2020).

Propagation-based approaches may work with a cascade depiction following a news propagation either based on hops, such as the number of reposts, or based on time factors where additional nodes are grouped in timeframes (Zhou & Zafarani, 2020). Zhou and Zafarani (2019, pp. 49–52) identified the following patterns: i) There are more people spreading fake news than there are spreading true news, ii) Propagation trees of fake news are deeper than those of true news, iii) Propagators show more engagement with fake news than with true news, iv) Fake news networks are more dense than true news networks. Even though they distinguish the mentioned sub-groups of context- and content-based approaches, Zhou and Zafarani (2020) advocate for combining the different approaches to benefit from the advantages of each approach. Liao et al. (2021) do so by using a pre-trained language model to gain insights into how news sentences relate to one another and to establish a heterogenous graph neural network where the nodes contain both the information from the text and attributes such as the number of followers or the published tweets.

Zhou and Zafarani (2020) emphasize the need for cross-domain fake news analysis, namely across different social networks, as a majority of the aforementioned research has been conducted with the same dataset which only contains data from Twitter (Liao et al., 2021; Ren & Zhang, 2021; Sitaula et al., 2020; Shu et al., 2019; Zhou & Zafarani, 2019; Yang et al., 2021; Shu, Silva, Wang, et al., 2017; Shu, Wang, & Liu, 2017). Wilson and Starbird (2020) report that for one specific news topic in the context of the Syrian civil war, the posts of users sharing an opposing view used significantly more cross-platform referencing than the users with a supporting view. Wilson and Starbird (2020) conclude that, to fully understand misinformation spreading, one needs to understand the complete courses of information-sharing across different online platforms.

Yang et al. (2021) advocated for cross-platform analysis of misinformation spreading because this depicts more the reality of today's information ecosystem, where information flows are simultaneous, multidirectional and exceeding the boundaries of a single platform. They confirmed that cross-platform studies were hardly conducted. This was mostly due to the challenge of evaluating data from heterogenous sources and the absence of a unified framework to make the different social media platforms comparable by showing their common attributes. Yang et al. (2021) approached these obstacles by identifying low-credibility sources on both Facebook and Twitter, thus obtaining comparable metrics for the two social media platforms. Different to Shao et al. (2018), Yang et al. (2021) found that social bots, or the automated spreading of misinformation, seemingly played a minor role when it came to COVID-19 misinformation spreading across Facebook and Twitter. However, they showed that specific accounts coordinated propagation of COVID-19 misinformation across the two different platforms.
Papakyriakopoulos et al. (2020) and Chen et al. (2021) both overcame the challenge of involving different social media platforms in a comparable manner by targeting URLs. With the keywords "COVID-19" and "corona", Papakyriakopoulos et al. (2020) extracted a large number of posts from a specific time period from different social media platforms. The posts were further queried by keywords relating to the origin of COVID-19. Then all third-party URLs were extracted from those posts and comments. Chen et al. (2021) analyzed the spreading of misinformation across different platforms by quantifying the spreading. They used the metrics total interaction and breakout scale as characterization of how a certain URL had been spread across different platforms. While total interaction referred to the number of likes, shares, comments, upvotes or retweets, depending on the platform, breakout scale categorizes a certain URL with a number from 1 to 3 according to the number of platforms on which the URL reached a certain threshold of total interaction (Chen et al., 2021). Papakyriakopoulos et al. (2020) found that misinformation spread more viral than neutral or clarifying content regarding COVID-19 conspiracy theories. Chen et al. (2021) also found that fake news URLs originating from different platforms had different probabilities of spreading across different platforms.

3 Method

The method of our research project can be divided into two phases: a major first part of our efforts went into the construction of a cross-platform dataset, comprising both fake news and trustworthy information, together with the corresponding "ground truth". In the second part, we then applied human-interpretable machine learning methods to discover multivariate spreading patterns in that data.

An overview of our approach is shown in **Figure 1**. We are using the two platforms Twitter and Reddit. The four depicted process steps can be described as follows:

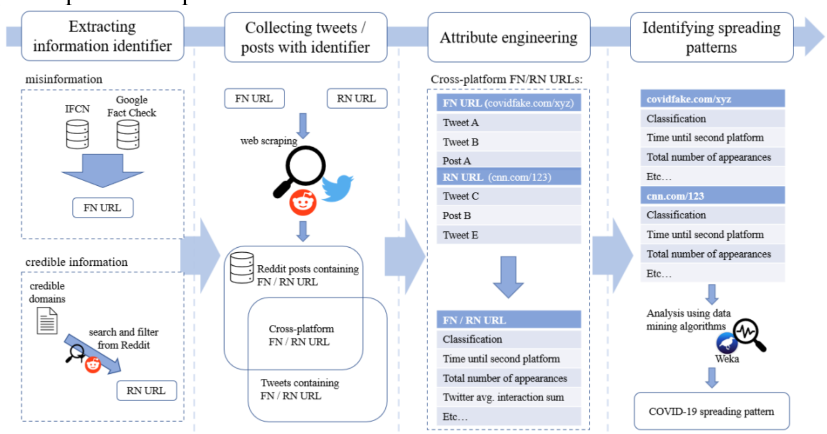

Figure 1: Overview of the research method

- **Collection of URLs:** we decided to identify "fake" or "real" social media posts on the basis of URLs contained in them. The advantage of URLs lies in the fact that they are unambiguous and that there are URL-based fact-checking datasets that we can use. Specifically, we worked with

  a) Corona-virus related fake news URLs (FN URLs) from the International Fact Checking Network (2022) and from the Google Fact Check Data Feed (2022)

  b) Corona-virus related "real news" URLs (RN URLs). These were obtained by using a list of "credible domains", as provided in (Mitchell et al., 2014; Pierri et al., 2020; Yang et al., 2021) and searching for URLs that a) stemmed from one of these domains and b) were used in Reddit posts near the keywords "Covid" or "Corona". We decided to extract full URLs from Reddit rather than Twitter because it represents the "data bottleneck".

- **Collection of posts:** we then worked on the assumption that posts containing FN URLs are "fake" and those with RN URLs are "real". We thus extracted all Reddit posts and Tweets containing any of the RN/FN URLs using Snscrape, a common web-scraping repository (JustAnotherArchivist, 2018). Our script looped through all RN/FN URLs using them as search terms backwards in time from the date of executing the scrape in April 2022 until the beginning of 2020. We defined a maximum of 1'000 instances per RN/FN URL. Then, we excluded all those where the contained URL did not occur in any entry in the other platform. That is, we ended up with a dataset of posts/tweets containing URLs that were used in both platforms. **Error! Reference source not found.** shows the number of URLs, posts etc. involved in these first two steps. As one may expect, there are more entries from Twitter than Reddit (roughly 50:1) and there is more credible information than misinformation (roughly 5:1).

- **Attribute engineering:** we decided to work with URLs as the data objects. That is, each URL was described with certain attributes derived from the Reddit posts and Tweets in which it occurred. These attributes were meant to describe the spreading process – they included e.g. the time that passed between the first occurrence of the URL on one platform and its first appearance on the other platform. Further attributes are described in more detail in Section 4.

- **Data mining:** Finally, we used the data mining suite WEKA (Frank et al., 2016) to apply supervised machine learning algorithms with human-interpretable models, capable of capturing multivariate patterns, specifically decision trees and rule learners, see Section 4 for details. The algorithms were given the attributes describing the spreading patterns of URLs and tasked to predict whether they were fake or real. We then studied the resulting models (trees and rule sets) to see what differentiates fake and real news regarding their spreading patterns.

Table 1: Quantitative overview of the data collection

Category	Quantity
FN URLs used as search terms for the scraping	1 '985
RN URLS used as search terms for the scraping	3'435
Scraped misinformation submissions and comments from Reddit	5'064
Scraped misinformation tweets	87'110
Scraped credible submissions and comments from Reddit	8'646
Scraped credible tweets	232'533
Reddit submissions and comments, including a cross-platform FN URL	3'335
Reddit submissions and comments, including a cross-platform RN URL	2'338
Tweets, including a cross-platform FN URL	46'497
Tweets, including a cross-platform RN URL	232'533
Unique cross-platform FN URLs	110
Unique cross-platform RN URLs	542

4 Findings and Discussion

In order to identify multivariate cross-platform spreading patterns of fake news, we used two different algorithms that are capable to produce human-readable multivariate patterns, namely a C.5 decision tree (Quinlan, 2014), called "J48" in Weka, as well as a rule learner based on Repeated Incremental Pruning (Cohen, 1995), called "JRip" in Weka. We chose these algorithms because they are a) globally (inherently) interpretable (Du et al., 2019), i.e. we can directly understand the complete set of patterns that their predictions are based on and b) as opposed to linear scoring systems (such as e.g. logistic regression, explained by Rudin et al. (2022)), they are capable of deriving multivariate patterns, i.e. ones that involve combinations of variables.

Each of the two algorithms was run over the data in 13 different configurations of attribute selection to cross-verify the results and get an optimal coverage of attribute variation. We defined a total of twelve attributes including the class attribute. Five of these are time related values such as the time difference between the first appearance on one social media platform and the first appearance on the other one or the entire timespan between the very first and the very last appearance. Three attributes are interaction and popularity indicators. Lastly, there is the class identifier, FN or RN, the unique identifier of each URL as well as the platform of first appearance, Twitter or Reddit.

```
Attributes:   9
              Classification
              Platform first appearance
              Time until second-platform in h
              Duration first to last appearance in h
              Total No of Appearances
              Avg time between appearances
              weighted avg time between appearances
              Twitter Total Interaction
              Twitter Avg Interaction Sum
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
------------------

Time until second-platform in h <= 2909.25
|   Duration first to last appearance in h <= 4946.35
|   |   Total No of Appearances <= 140: Credible Information (205.0/14.0)
|   |   Total No of Appearances > 140
|   |   |   Time until second-platform in h <= 26.37
|   |   |   |   Duration first to last appearance in h <= 1506.7: Misinformation (10.0)
|   |   |   |   Duration first to last appearance in h > 1506.7
|   |   |   |   |   Twitter Total Interaction <= 574: Misinformation (2.0)
|   |   |   |   |   Twitter Total Interaction > 574
|   |   |   |   |   |   Total No of Appearances <= 470: Credible Information (4.0)
|   |   |   |   |   |   Total No of Appearances > 470: Misinformation (6.0/2.0)
|   |   |   Time until second-platform in h > 26.37: Credible Information (53.0/6.0)
|   Duration first to last appearance in h > 4946.35
|   |   Time until second-platform in h <= 7.8: Misinformation (29.0/4.0)
|   |   Time until second-platform in h > 7.8
|   |   |   Twitter Avg Interaction Sum <= 3.504348
|   |   |   |   Avg time between appearances <= 107.258705: Misinformation (35.0/8.0)
|   |   |   |   Avg time between appearances > 107.258705: Credible Information (8.0/1.0)
|   |   |   Twitter Avg Interaction Sum > 3.504348: Credible Information (109.0/20.0)
Time until second-platform in h > 2909.25: Credible Information (191.0/1.0)

Number of Leaves  :     11

Size of the tree :      21
```

Figure 2: J48 decision tree from Run 1

We chose an experimentative approach, varying with those attributes that seemed to have the most impact on the decision tree in order to see if new patterns emerged when the dominating attributes of the formerly explored patterns were removed. Figure 2 and Figure 3 show the results of such an algorithm run on our dataset with J48 and JRIP respectively. Table 2 shows an overview of the five identified main patterns and their appearances in both the J48 decision tree's and the JRIP rule learner's results.

Table 2: Overview of patterns and their appearance

Pattern	**J48 runs confirming the pattern**	**JRIP runs confirming the pattern**
"Time until second platform in h" > 2909.25: Credible Information (191.0/1.0)	Run 1 Run 8 Run 2 Run 9 Run 3 Run 11 Run 4 Run 12 Run 5 Run 13 Run 6 Run 7	The following runs included the attribute with the same threshold value of 2909.25 hours: Run 1 Run 7 Run 4 Run 12 Run 8 included the attribute with a close value of 2089.98 in one of its rules.
"Duration first to last appearance in h" <= 4 946.35 and "Total No of Appearances" <= 140: Credible Information (205.0/14.0)	Run 1 Run 8 Run 6[*] Run 9[*] Run 2[*] Run 11[†] Run 3[*] Run 4[*] Run 5[*] Run 7	The following runs included the attribute "Duration first to last appearance in h" with a similar threshold of 4 948.25: Run 1 Run 8 Run 7 Run 3 Run 2 Run 13

[*] This run had the additional rule "Platform of first appearance: Twitter" prior to the described pattern. This led to the following alternative leaf at the end of the pattern: "Credible Information (195.0/12.0)".

[†] This run showed a similar pattern with different threshold values. It was still considered as confirming of the pattern.

Always showing predominantly misinformation instances below the threshold of "Time until second platform in h"	Run 1 Run 4 Run 7 Run 11 Run 2 Run 5 Run 8 Run 12 Run 3 Run 6 Run 9 Run 13	Run 1 Run 4 Run 7 Run 11 Run 2 Run 5 Run 8 Run 12 Run 3 Run 6 Run 9 Run 13
Platform first appearance = Reddit and Total No. of Appearances > 1 008: Misinformation (14.0/1.0)	Run 2 Run 3 Run 4 Run 5 Run 6 Run 9	Run 4 included the following confirming rule: (Time until second platform in h <= 2 909.25) and (Platform first appearance = Reddit) and (Total No. of Appearances > =1009) =>Classification=Misinformation (12.0/1.0) =>Classification=Credible Information (558.0/44.0)
"Twitter Avg Interaction Sum" between 3.4 and 3.9 as a threshold in combination with a "Total No. of Appearances" threshold of around 150	Run 10 Run 12	"Twitter Avg Interaction Sum" threshold between 3.4 and 3.9, and predominantly misinformation below that threshold: Run 1 Run 10 Run 5 Run 11 Run 7 Run 12 Run 8 Run 13

The clearest of the spreading patterns showed that information that takes more than 121 days (2909.25 h) is mostly credible information. Furthermore, the attribute "Time until second platform" seemed to group the data in most runs. Repeatedly, thresholds for this attribute were chosen by the algorithms around 8, 26, 46 and 60 hours. Remarkably, in every single algorithm run conducted, the models consistently showed misinformation dominantly appearing below the threshold, indicating that misinformation has a tendency of spreading faster from one social media platform to the other.
This might be explained with an intentional process behind cross-platform misinformation spreading. Misinformation might be deliberately copied quicker from one platform to another and propagated more often in the interest of reaching as many readers as fast as possible and maybe gain credibility through accumulation. This would be in accordance with the findings of Yang et al. (2021) who presented evidence of coordinated misinformation sharing on two social media platforms.

```
Attributes:   9
              Classification
              Platform first appearance
              Time until second-platform in h
              Duration first to last appearance in h
              Total No of Appearances
              Avg time between appearances
              weighted avg time between appearances
              Twitter Total Interaction
              Twitter Avg Interaction Sum
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

JRIP rules:
===========

(Time until second-platform in h <= 60.13) and (Total No of Appearances >= 154) and (Twitter Avg Interaction Sum <= 3.414847) =>
Classification=Misinformation (20.0/3.0)
(weighted avg time between appearances >= 27.485166) and (Duration first to last appearance in h >= 4948.25) and (Time until second-platform in h <=
38.37) => Classification=Misinformation (32.0/8.0)
(Time until second-platform in h <= 2909.25) and (Twitter Avg Interaction Sum <= 3.453488) and (Total No of Appearances >= 151) =>
Classification=Misinformation (21.0/5.0)
(Total No of Appearances >= 147) and (Time until second-platform in h <= 8.23) => Classification=Misinformation (17.0/3.0)
 => Classification=Credible Information (562.0/39.0)

Number of Rules : 5
```

Figure 3: JRIP rules from Run 1

Another clear pattern, which is shown in Figure 2 as well as Figure 3, is that information that reappeared for less than 206 days (Duration first to last appearance), while also being less popular (Total No of Appearances <= 140) was mostly credible information. In fact, a threshold somewhere between 140 and 150 number of appearances often differentiated credible information from misinformation in our results, with misinformation mostly being dominant above the threshold. Hence, it can be said that misinformation seems to spread more virally than credible information, while also having a tendency of reappearing more often and over a longer period. This matches the findings of Vosoughi et al. (2018) and Zhao et al. (2020) who observed misinformation also to spread more virally on one platform. Furthermore, it fits in with Papakyriakopoulos et al. (2020) who found misinformation to spread more virally on different social media platforms. Our finding also seems to correspond with Yang et al.'s . (2021) conclusion of credible information having a lower prevalence than misinformation. We might speculate that credible information wears out faster or that the mentioned intentional propagation process deliberately keeps a misinformation alive e.g., cares for multiple reappearances over a longer period.

Our results indicate which attributes are more effective in distinguishing credible information and misinformation from each other based on their spreading patterns. The following features distinguished the most instances: i) "Time until second platform", appearing in all runs the attribute was included, ii) "Total number of appearances", appearing in all runs the attribute was included, iii) "Duration of appearance", appearing in eleven runs of each algorithm and iv) "Twitter Average Interaction Sum", appearing in seven J48 runs and eleven JRip runs. Seemingly the platform of origin played a minor role in misinformation spreading in our data set. Differently than Chen et al. (2021), we found that once misinformation spreads across different platforms, the exact platform of origin could be disregarded as a pattern defining factor. They stated that fake news URLs originating from different platforms had different probabilities of spreading across different platforms. The only clear pattern we found here indicated that information originating on Reddit and being rather popular with appearances beyond 1008 times was mostly misinformation. Since Reddit generally shows less volume of appearances than Twitter due to its smaller user base, this finding can again be interpreted as a generally bigger push on volume for misinformation.

There was a specific threshold throughout the different algorithm runs, often between 3.4 and 3.9, for the attribute "Twitter Average Interaction Sum". The instances below that threshold were mostly misinformation, whereas the ones above the threshold were mostly credible information. Therefore, if an instance was fairly popular with more than 145 appearances but not that much interaction took place with each of those appearances, it is more likely misinformation. This is aligned with Zhou & Zafarani's (2019) findings of propagators showing more engagement with fake news than with true news. It also raises the question of how information became that fairly popular, if seemingly people did not interact with it so much. The most obvious explanation would be that the spreading was pushed intentionally and did not simply follow chance.

Apart from these findings regarding the importance of certain attributes, our discovered are truly multivariate, in the sense that both the tree and the rules always combine multiple attributes, e.g. the first rule in Figure 3 refers to a short time of spreading from one platform to the other, *in conjunction* with a large number of appearances, *but* showing relatively little interaction. This shows that patterns can be more complex than the isolated metrics used in previous work.

While our findings showed that misinformation spreads faster across social media platforms, there was no pattern visible that included the two attributes related to time. These attributes bear information regarding the time in between single appearances of misinformation and credible information. These attributes were originally intended as indicators for spreading speed. While these two attributes hardly found application in the J48 analysis, they did appear several times using JRIP. However, no clear tendency was visible.

5 Conclusions

In this work, we have, for the first time, used a cross-platform dataset to discover multivariate spreading patterns of fake news *across* more than one social media platform. Our analysis was based on a collection of both real and fake news URLs who were referenced in the two platforms Twitter and Reddit and whose spreading processes – above all the process of spreading from one platform to the other – we have described with a number of features. By applying supervised machine learning with human-interpretable classifier models, we were able to elicit a few multivariate patterns in the form of decision trees and rules. Some of these patterns were observed repeatedly across multiple configurations and learning algorithms. The clearest pattern that the data has indicated is that information that takes more than 4 months to spread from one platform to another tends to be credible information or more generally – with even more evidence supporting it – that information spreading faster to other platforms tended to be misinformation. Another significant finding was that, once misinformation spread across different social media platforms, the platform of origin had a minor influence on spreading patterns. Finally, there was a prevalence of misinformation appearing with a rather high frequency on both platforms even though their interaction rate on Twitter is rather low.

A possible interpretation of these findings is of course that cross-platform spreading of fake news is an intentional process, i.e. the originators or propagators of misinformation might have an interest in pushing both the speed and the volume of the spreading on as many platforms as possible.

In terms of future work, other researchers are encouraged to use and build on the dataset and enrich it with more cross-platform data. As mentioned before, the collection of misinformation data and general data availability remain a bottleneck in this field of research. With our dataset, we hope to have made a first step in the direction of developing a cross-platform dataset that can inform a larger body of research. Specifically, future research may benefit from adding further platforms, e.g. Facebook, and making more attributes available that the spreading process of both fake and real news instances can be described with and to investigate their relative usefulness.

References

Chen, Z., Aslett, K., Reynolds, J. R., Freire, J., Tucker, J. A., & Bonneau, R. (2021, April 15). *An Automatic Framework to Continuously Monitor Multi-Platform Information Spread* [CEUR Workshop Proceedings]. CEUR Workshop. http://ceur-ws.org/Vol-2890/paper3.pdf

Cohen, W. W. (1995). Repeated incremental pruning to produce error reduction. *Machine Learning Proceedings of the Twelfth International Conference ML95*.

Data Commons. (2022). *Fact Check Markup Tool Data Feed*. Fact Check Markup Tool Data Feed. https://storage.googleapis.com/datacommons-feeds/claimreview/latest/data.json

Du, M., Liu, N., & Hu, X. (2019). Techniques for interpretable machine learning. *Communications of the ACM*, *63*(1), 68–77.

Frank, E., Mark A. Hall, & Witten, I. H. (2016). *The WEKA Workbench* (4th ed.). Morgan Kaufmann. https://www.cs.waikato.ac.nz/ml/weka/

Haron, H. H. B. (2022). Infodemic Laws in Malaysia: Lessons, Challenges, and Suggestions for the Future in Light of the Covid-19 Pandemic2. *Malaysian Society 5.0: Selected Contemporary Socio-Legal Issues*, 31.

International Fact-Checking Network. (2022). IFCN Covid-19 Misinformation. *Poynter*. https://www.poynter.org/ifcn-covid-19-misinformation/

JustAnotherArchivist. (2018). *Snscrape* [Python]. https://github.com/JustAnotherArchivist/snscrape

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., & others. (2018). The science of fake news. *Science*, *359*(6380), 1094–1096.

Liao, H., Liu, Q., Shu, K., & xie, X. (2021). Fake News Detection through Graph Comment Advanced Learning. *ArXiv:2011.01579 [Cs]*. http://arxiv.org/abs/2011.01579

Mitchell, A., Gottfried, J., Kiley, J., & Matsa, K. E. (2014, October 21). Political Polarization & Media Habits. *Pew Research Center's Journalism Project*. https://www.pewresearch.org/journalism/2014/10/21/political-polarization-media-habits/

Oshikawa, R., Qian, J., & Wang, W. Y. (2018). A survey on natural language processing for fake news detection. *ArXiv Preprint ArXiv:1811.00770*.

Papakyriakopoulos, O., Serrano, J. C. M., & Hegelich, S. (2020). The spread of COVID-19 conspiracy theories on social media and the effect of content moderation. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-034

Parikh, S. B., & Atrey, P. K. (2018). Media-rich fake news detection: A survey. *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, 436–441.

Pierri, F., Piccardi, C., & Ceri, S. (2020). A multi-layer approach to disinformation detection in US and Italian news spreading on Twitter. *EPJ Data Science*, *9*(1), Article 1. https://doi.org/10.1140/epjds/s13688-020-00253-8

Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., & Stein, B. (2017). A Stylometric Inquiry into Hyperpartisan and Fake News. *ArXiv:1702.05638 [Cs]*. http://arxiv.org/abs/1702.05638

Quinlan, J. R. (2014). *C4. 5: Programs for machine learning*. Elsevier.

Ren, Y., & Zhang, J. (2021). Fake News Detection on News-Oriented Heterogeneous Information Networks through Hierarchical Graph Attention. *2021 International Joint Conference on Neural Networks (IJCNN)*, 1–8. https://doi.org/10.1109/IJCNN52387.2021.9534362

Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L., & Zhong, C. (2022). Interpretable machine learning: Fundamental principles and 10 grand challenges. *Statistic Surveys*, *16*, 1–85.

Shao, C., Ciampaglia, G., Varol, O., Flammini, A., Menczer, F., & Yang, K.-C. (2018). The spread of low-credibility content by social bots. *Nature Communications*, *9*. https://doi.org/10.1038/s41467-018-06930-7

Shu, K., Silva, 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. https://doi.org/10.1145/3137597.3137600

Shu, K., Wang, S., & Liu, H. (2017). *Exploiting Tri-Relationship for Fake News Detection*.

Shu, K., Wang, S., & Liu, H. (2019). Beyond News Contents: The Role of Social Context for Fake News Detection. *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, 312–320. https://doi.org/10.1145/3289600.3290994

Shu, K., Wang, S., Tang, J., Zafarani, R., & Liu, H. (2017). User identity linkage across online social networks: A review. *Acm Sigkdd Explorations Newsletter*, *18*(2), 5–17.

Sitaula, N., Mohan, C. K., Grygiel, J., Zhou, X., & Zafarani, R. (2020). Credibility-Based Fake News Detection. In K. Shu, S. Wang, D. Lee, & H. Liu (Eds.), *Disinformation, Misinformation, and Fake News in Social Media: Emerging Research Challenges and Opportunities* (pp. 163–182). Springer International Publishing. https://doi.org/10.1007/978-3-030-42699-6_9

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Wilson, T., & Starbird, K. (2020). Cross-platform disinformation campaigns: Lessons learned and next steps. *Harvard Kennedy School Misinformation Review*, *1*(1). https://doi.org/10.37016/mr-2020-002

Yang, K.-C., Pierri, F., Hui, P.-M., Axelrod, D., Torres-Lugo, C., Bryden, J., & Menczer, F. (2021). The COVID-19 Infodemic: Twitter versus Facebook. *Big Data & Society*, *8*(1), Article 1. https://doi.org/10.1177/20539517211013861

Zhang, X., & Ghorbani, A. A. (2020). An overview of online fake news: Characterization, detection, and discussion. *Information Processing & Management*, *57*(2), Article 2. https://doi.org/10.1016/j.ipm.2019.03.004

Zhao, Z., Zhao, J., Sano, Y., Levy, O., Takayasu, H., Takayasu, M., Li, D., Wu, J., & Havlin, S. (2020). Fake news propagates differently from real news even at early stages of spreading. *EPJ Data Science*, *9*(1), 7. https://doi.org/10.1140/epjds/s13688-020-00224-z

Zhou, X., & Zafarani, R. (2019). Network-based Fake News Detection: A Pattern-driven Approach. *ACM SIGKDD Explorations Newsletter*, *21*(2), 48–60. https://doi.org/10.1145/3373464.3373473

Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, *53*(5), 1–40.