Real-Time Prediction of Online False Information Purveyors and their Characteristics

Anil R. Doshi[*], Sharat Raghavan[†], William Schmidt[‡]

October 30, 2020

Abstract

Disinformation, misinformation, and other 'fake news'—collectively false information—is quick and inexpensive to create and distribute in our increasingly digital and connected world. Identifying false information early and cost effectively can offset some of those operational advantages. In this paper, we develop light-weight machine learning models that utilize (1) a novel data set tracking browsing behavior and (2) domain registration data that is available for all websites when they are established. Using only the domain registration data, we develop and demonstrate a machine learning classifier that identifies domains, at the time the domain is registered, that will go on to produce false information. We then combine this data with our browsing data and develop a machine learning classifier that identifies false information domains whose content is most associated with higher levels of consumption. Finally, we use our data to identify false information domains that will cease operations after an event of interest, in our case the 2016 U.S. presidential election. We theorize that the last category involves actors seeking primarily to manipulate perceptions and outcomes of that event.

1 Introduction

The online proliferation of disinformation, misinformation, and other 'fake news'—collectively false information—has become an increasingly common characteristic of the digital information environment Bradshaw and Howard (2019). Recent false information campaigns have targeted areas that are salient to management and operations in both the private and public sectors. Some false information campaigns target companies. For example, the United States Department of Homeland Security identified a false information campaign in 2018 in which "right wing actors . . . sought to discredit and undermine Nike's brand

[*] UCL School of Management, Level 38, One Canada Square, London E14 5AA, UK. anil.doshi@ucl.ac.uk. Corresponding author.

[†] Haas School of Business, University of California at Berkeley.

[‡] Cornell University SC Johnson School of Business; wschmidt@cornell.edu. We thank DomainTools for sharing domain data with us and Hunt Allcott and Matthew Gentzkow for generously sharing their data with us.

reputation" and "do economic harm to a corporation with whom they disagreed" (U.S. Department of Homeland Security, 2019). In 2020, Facebook accused one of south-east Asia's biggest telecommunication firms of using Facebook accounts to conduct a commercial disinformation campaign seeking to discredit its competitors (Murphy and Reed, 2020). False information campaigns can also target local and national communities and governments. For instance, in 2014, elaborately orchestrated false information campaigns separately fabricated an explosion at a chemical plant in Louisiana and an outbreak of the Ebola virus in Atlanta, Georgia (Chen, 2015). And in 2020, the United States accused Russia, China, and Iran of engaging in far-reaching false information campaigns on the causes, treatments, and consequences of the novel COVID-19 pandemic (Barnes et al., 2020).

In this paper, we use data from a canonical example of organized false information – the 2016 U.S. presidential election. We show how light-weight machine learning models that utilize data at the time a website is registered can be used as an early warning signal to identify domains that are likely to (i) produce false information, (ii) produce false information that is most associated with high levels of consumption, and (iii) abandon their operations after an event of interest, in our case an election. We theorize and provide suggestive evidence that domains abandoned after an event of interest are established by entities seeking primarily to manipulate perceptions and outcomes of that event.

The prevalence of false information activity is well established in our setting. For instance, researchers tracked a sample of 156 fake news stories distributed during the 2016 presidential election and found they were shared on social media over 37.6 million times in the months leading up to the 2016 U.S. election (Allcott and Gentzkow, 2017). Importantly, domain experts assert that successful methodologies from false information campaigns in election settings will be utilized and adapted to other settings. The *Financial Times* quoted Nathaniel Gleicher, head of Facebook's cyber security policy, as saying, "We would expect other types of actors, such as corporate actors, to see these types of [electoral] disinformation campaigns and follow that model" (Murphy and Reed, 2020). As a result, we believe that our findings in this setting can also shed useful insights into other non-election settings.

False information detection is an active research area for many scholars and practitioners. Scholars have sought to shed light on motivations for false and deceptive information, including insights on why companies would post fake reviews on Yelp (Luca and Zervas, 2016) and the impact of social media influencers on information aggregation and the propagation of deceptive information (Acemoglu et al., 2010). Other studies examine how false information is promoted and shared on social networks (Del Vicario et al., 2016; Vosoughi et al., 2018; Grinberg et al., 2019). Emerging work examines operational policies that can inform the how social media platforms respond to false information (Candogan and Drakopoulos, 2020; Papanastasiou, 2020).

The first automated efforts to prevent the spread of disinformation largely focused on linguistic characteristics of online content and the network characteristics of social media proliferation (Conroy et al., 2015). Subsequent research has employed increasingly complex textual models and user characteristics to identify false information (Shu et al., 2017; Castelo et al., 2019). This prior work almost exclusively uses article text or social media content as the source of features in a variety of models. Our classifiers, on the other hand, rely only on seemingly benign information that is available at the time the domain is registered, and is required of every domain on the internet. This means that it can be implemented earlier than existing approaches, and even prior to the appearance of content on the domain.

Purveyors of false information are shifting away from efforts dominated by social media, and are increasingly relying on the production and provision of false information articles on website domains that appear to be legitimate news outlets. This can "exploit the credibility of local journalism" (Coppins, 2020), while still employing social media as a means to promote the false information. *The New York Times* cites experts who indicate that this trend is fed by the realization that false information articles and the sites that host them are more difficult to identify, and thus, combat (Rosenberg and Barnes, 2020). Websites are relatively easy to set up and costless to abandon, allowing the entities behind them to be nimble and avoid countermeasures. And once false information is seeded into the digital information environment, websites can rely on independent individuals with like-minded objectives to disseminate links to that information on their social media networks. All of this motivates our focus on false information websites, and our effort to develop a model that can be used to identify false information domains early in their life cycle, before they have achieved their objectives.

Ferreting out false information is an asymmetric game that favors producers of content over those seeking to mitigate it. The amount of effort required to combat online false information using traditional text classification tools is high, due in part to the massive volume of text produced every day on the web. This contrasts sharply with the relatively low amount of effort required to develop and publish false information. Trend Micro, a cybersecurity company, analyzed scores of false information providers and report that such services can be deployed quickly and cheaply—false information websites can be purchased for about \$3,000 and maintained with relevant fake content at a cost of \$5,000 per month (Gu et al., 2017). The existence of low-cost, agile false information providers reinforces our objective to help counter them by developing a scalable and computationally efficient model.

We construct three classifiers using two retrospective samples—domains that were identified as being purveyors of false information and a random selection of contemporaneously established domains. We train and test our classifiers using data drawn from the registration of each domain with the International Corporation for Assigned Names and Numbers (ICANN), the release date of false information articles, and the browsing history of a sample of U.S. internet users.

Our first model relies exclusively on domain registration data to predict whether domains will become purveyors of false information or not. Our approach is inspired by Guzman and Stern (2015), who use company characteristics available at the time an entrepreneur registers her startup to predict which startups will be successful. Our second model predicts an outcome based on false-information consumption derived from a novel dataset of user browsing activity leading up to the 2016 U.S. election. Our third model identifies false information providers with a certain operating profile, which we argue may be indicative of the domain's objective. In particular, we predict which false information domains will cease operations shortly after the election, and theorize that this is a proxy for domains whose primary objective was to manipulate the news environment or voter behavior in the run-up to the election.

To summarize our results, we provide the specificity (true positive rate) and sensitivity (true negative rate) of each of the models using a cut point of 0.7.[1] We also provide the area under the curve (AUC), a measure used to assess the performance of prediction models with binary outcomes. Our classification of whether a domain is a false information provider yields values for specificity and sensitivity of 92.0% and 96.2%, respectively, and the AUC is 0.85. Our classification of whether a false information provider's content is associated with higher browsing activity yields values for specificity and sensitivity of 94.1%, and 80.0%, and the AUC is 0.84. Finally, our classification of whether a false information provider discontinues operations shortly after the 2016 election yields values for specificity and sensitivity of 88.2% and 84.4%, and the AUC is 0.69.

Our early-identification system can help policy makers deploy their limited resources more rapidly and effectively by prioritizing domains for potential sanction or increased monitoring. In this way, our approach can complement other, heavier-weight machine learning models, such as text-based classifiers. By using our early-identification system in conjunction with a staged escalation process and other validation tools, policy makers can mitigate the possibility of taking action based on possible false positive classifications, which are inherent in any machine learning system.

We make several contributions to the growing work around false information prediction. First, by focusing on the website domains rather than social media, we are addressing a resurgent and challenging front of false information campaigns. Second, our models use data that is available at the time a domain is registered, thereby allowing for early identification of false information domains before they can get established. This data is required and well-structured for any registered domain, which facilitates easier access and use. Finally, our models employ machine learning tools that are efficient and scalable, which facilitates a rapid and potentially automated analysis. These contributions will help institutions that are grappling with identifying and responding to false information, including regulatory agencies, credible news providers, and technology and platform firms.

[1] Changing the false positive and false negative rates can be accomplished by changing the cut point.

2 Data

Our analysis exploits a novel database provided by our research partner, the Mozilla Corporation, developer of the Firefox browser. Mozilla recruited U.S. Firefox users to participate in an unrelated study during which their web browsing habits were monitored for a period of time that coincided with the run-up to the 2016 election. On the eve of the election, there were 2,680 users participating in the study. In the 30 weekdays prior to the election, those participants had collectively visited 2,670,124 webpages. Of those visits, 26,310 (1.0% of the total) were to false information sites. Mozilla provided us with data on daily browsing activity classified by the following domain types—false information, credible news, social media,[2] and all other.

We identify false information domains and content using a database provided by Allcott and Gentzkow (2017), who aggregated articles that were confirmed as being false by at least one of the following fact-checking services: Snopes, Politifact, or Buzzfeed. The authors describe the database as "a reasonable but probably not comprehensive sample of the major fake news stories that circulated before the election." (Allcott and Gentzkow (2017); p. 219). After dropping articles that did not contain any domain information or were registered after Election Day, we are left with a sample of 883 false information articles hosted on 363 domains.

Our sample of general domains registered prior to the election was provided by DomainTools, an online security company. The details of the sample generation process are provided in the Online Appendix. DomainTools also furnished us with the domain registration information for all the domains in our sample. To obtain a website domain, one must register it with ICANN, the nonprofit that manages domains. A user who registers a domain provides certain information, including the name of the domain and its extension (e.g. ".com" or ".org") as well as the names and contact information for the registrant and site, billing, and technical administrators. The registration date is also recorded and attached to each record.

[2] Social media domains consisted Facebook, Twitter, LinkedIn, Instagram, Snapchat, and Pinterest. These services represented approximately 94.3% of all social media visits in 2016 StatCounter (2016).

2.1 Outcome Measures

We use three outcomes in our analyses. Our first outcome is whether a site is a false information domain. This is generated using the observations identified in Section 2, and consists of 363 false information domains and 1,861 general domains.

Our second outcome is based on the relative efficacy of each false information domain. For each domain, we calculate the average number of visits captured by our Mozilla data to *all* false information domains on days when the focal false information domain publishes an article that appears in the Allcott and Gentzkow (2017) database.[3] For example, if a domain publishes false information articles on two days in our study that have 19 and 15 visits to false information domains by the participants in our data, respectively, then the measure for this domain is 17 ((19+15)/2=17). For our second outcome, we identify whether the resulting value for a domain is larger than the median value across all false information domains.

Our third outcome is based on whether the domain discontinued operations as of June 2017, approximately seven months after the election. We determined this by visiting each domain If the domain could not be resolved by the browser or there was some placeholder landing page with text like "Buy this domain," we concluded that the site had ceased operations. We found that 96 of the domains (27.0%) met this criterion.

2.2 Feature Extraction

To create each of our prediction models, we generated a set of features from the domain registration data. Specifically, we include a binary variable that equals "1" if there is an individual or institutional name in the billing contact field. We add categorical variables for the registrant name being an individual, an institution, or private, We include additional categorical variables for the domain extension, registrar, registration state, and country. To utilize the domain name information, we create a set of binary variables that equal "1" when the domain included the following terms that pertain to the election: *trump*,

[3] Our measure accounts for visits to all false information domains because it is common practice for articles that break on one domain to receive coverage and even full reproduction on other domains. Similar practices can be found in credible news outlets (Boczkowski, 2010), albeit not full reproduction of content.

conservative, *clinton*, and *liberal*. We also create several date-related features, including the registration date, year, year-squared, and individual year binary variables. Finally, we create interactions between each of the year, year-squared, and 2016 indicators and all the other non-date based features. This yields a total of 957 features[4]

3 Methods

To build our models, we employ a technique of variable selection and model fitting from machine learning called LASSO (Friedman et al., 2010). LASSO is a form of penalized regression, which effectively adds a penalty term to the standard ordinary least squares (OLS) optimization problem that moves weights on features toward zero. The penalty is imposed to avoid the problem of overfitting, which occurs when the model fits the given data well, but then performs poorly out of sample.

To assess the performance of the model, we employ 10-fold cross validation. Each model is fit ten times, with each iteration using 90% of the total sample to fit the model. The final 10% of holdout data is used to assess the performance of the model. Each model yields a prediction of the outcome that varies from zero to one.

4 Results

Each of our classifier models provides a predicted probability of whether the domain is classified as the outcome and ranges from zero to one. From this continuous measure, we can set a cut off threshold, and then assess the specificity (true positive rate) and sensitivity (true negative rate) of the model. We describe the model performance at a threshold of 0.7; that is, if the continuous predicted probability is greater than or equal to 0.7, we classify it as being labelled by the appropriate outcome. For values less than 0.7, we classify it as not being labelled with the appropriate outcome. For example, in our first model, we classify a domain as being a false information producers if its continuous predicted value is greater than or equal to 0.7. If the value is less than 0.7, we classify it as not being a

[4] Note that the number of features in each analysis differ for two reasons. First, we remove any date-related features from our false information versus credible news model. Second, the samples differ across each analysis.

false information producer. A higher value for the cutoff will result in a higher levels of specificity, at the expense of lower levels of sensitivity.

For each model, we also depict the receiver operating characteristic (ROC) curves. The ROC curve visually depicts the tradeoffs between the true positive rate and the false positive rate (or sensitivity and one minus the sensitivity). Each point on the curve represents the two performance metrics for a chosen threshold. From the ROC curve, we compute the area under the curve (AUC) metric used to assess model performance.

We conceive that, in practice, the first model can act as a first line of defense that provides an early warning that identifies likely false information providers. The analysis would conceivably lead to the flagging of suspect domains for intervention, further data collection, or escalated monitoring. Our next two analyses would be applied to domains that were subsequently confirmed to contain false information (perhaps by using human readers or a text-based machine learning algorithm), and would assist with the deployment of more robust policy responses.

4.1 Distinguishing False Information Domains

Our first model distinguishes which domains are purveyors of false information. This model is applied against the full sample of 363 false information domains and 1,861 general domains. We exclude date-related variables, in the event those variables would unduly advantage our prediction model. Our results are slightly stronger if we include these variables. The model includes 500 features and assigns non-zero weights whose absolute value exceeds 0.001 to 151 features. As depicted in Figure 1, the AUC achieved by the model is 0.85. Using the 0.7 threshold, the model achieves specificity and sensitivity of 92.0% and 96.2%, respectively.

4.2 False Information Production on High Consumption Days

In our second analysis, we assess whether the level of false-information consumption in the future varies by prior observable characteristics of false-information producers. We motivate this question with insights from a model-free analysis from the six week period prior to the 2016 election. Using the 30 weekdays over that period (due to differences in creation and consumption of news between weekdays and weekends, in general as per Boczkowski (2010)), we compile the number of visits per participant on each day (because the size of the sample is not fixed) and use it as a proxy for attentiveness to false information on that day. For each article published on a domain in our sample, we calculate the percentage change in false-information consumption as the amount of false information consumed on the publication date relative to the average of the amount of false information consumed in the five surrounding weekdays. We can collapse this to a domain-level measure by averaging the false information domain visits over all days that the producer generated a false information article. Figure 2 is a sorted bar chart of this measure for 201 domains in our study window.[5] As expected, the change in false-information consumption is predominantly positive when there is false-information production. Less expected, however, is the wide variation in the percentage change in false-information consumption across producer-days. The minimum of the percentage change in false-information consumption is -31.7% and the maximum is 49.3%. The goal of our second model is to use domain registration data to predict which false information domains produce articles on these higher consumption days.

Our sample includes the 201 domains that published at least one false information article during our study period. The domain registration dataset includes 431 features (including an intercept), and our LASSO model assigns a non-negative weight to 188 features, of which 113 have an absolute value greater than 0.001. Figure 3a shows the ROC curve for this model (AUC = 0.84). At the 0.7 threshold, specificity and sensitivity are 94.1% and 80.0%, respectively.

4.3 Domains that Discontinued Operations

Our next model is motivated by the observation that domains which discontinued their operations shortly after the election were actually more successful at inducing browsing activity before the election compared to domains that continued operations. To investigate this issue, we manually check whether domains were still active as of June 2017. The

[5] We note that the remaining false information domains from our sample do not appear in this analysis, because they did not publish articles in the 30 weekdays period prior to the election.

timing of our search allows for registration deactivation to take place when owners fail to renew contracts that typically run annually or on a multi-year basis. We find that 96 of the 363 false information domains in the dataset (27.0%) were no longer active. Using this information, we categorize domains based on whether they discontinued operations shortly after the election or remained an ongoing concern.

To highlight the operational effectiveness of discontinued domains, we include this categorization in a regression to estimate false-information consumption. The results are presented in Table 1. In this regression analysis, *Discontinued Production* captures the number of false information articles produced on each day by domains that discontinued their operations shortly after the election. *Ongoing Production* captures the number of false information articles produced on each day by domains that maintained their operations. *Social Media Visits* controls for the number of visits to social media domains. While it might be expected that false-information producers that ceased operations were associated with lower levels of false-information consumption, we find the opposite is true. The difference between the coefficient on *Discontinued Production* and the coefficient on *Ongoing Production* is *positive* and statistically significant. This comparison is tested using a Chi-square test whose test statistic values are presented in the table. This finding does not conform to prior categorizations of false-information producers as being motivated by economic or ideological objectives (Allcott and Gentzkow, 2017).[6] In the Discussion section, we argue that this operational characteristic may identify producers that are instead motivated to manipulate an event of interest, and therefore abandon their operations shortly after the conclusion of the event.

From a policy perspective, such an ex-post identification of sites that are discontinued does not provide any actionable way of identifying which sites to sanction in advance. As a result, we present a third prediction model, in which we predict whether the site will eventually discontinue its operations, using all 363 false information domains from Allcott and Gentzkow (2017). The model was performed on 431 features; LASSO assigned non-zero weights to 144 features, of which 95 were assigned an absolute weight greater than 0.001.

[6] Economic actors are those with traditional financial incentives in the production of any good, and ideological actors are those that produce content to promote a specific viewpoint. These types need not be mutually exclusive.

Figure 3b presents the ROC curve for this model (AUC = 0.69). At the 0.7 threshold, specificity and sensitivity are 88.2% and 84.4%, respectively.

5 Discussion

In this paper, we demonstrate that it is possible to develop machine learning models that generate high quality classifications of false information providers based on immediately available information, specifically data furnished at the time a domain is first registered with regulatory agencies. False information providers have the ability to enter and leave the marketplace quickly, and enjoy a low cost of production and operation. By developing a light-weight, scalable model that relies on data that is available at the time a domain is registered, we can help to counteract these advantages. Doing so is important because false information affects our real, non-digital environment, including managerial decision making, customer perceptions, election outcomes, and a host of personal and organizational choices.

Our analyses may have broader applications beyond those that we explicitly identify. Allcott and Gentzkow (2017) indicate that fake news producers are motivated by economic (an interest in generating revenue, regardless of content type) or ideological (an interest in promoting a particular viewpoint) considerations. A related, but potentially unique motivation is to manipulate the outcome of a particular event. Our third model, in which we categorize producers based on whether they discontinued operations shortly after the election, could be redeployed to identify false information providers with such motivations. This is motivated by the observation that articles from discontinued domains were associated with significantly more visits to false information domains prior to the election, raising the question of why those false information providers would shut down a seemingly effective domain once the election ended. We think such operational behavior could serve as a proxy for domains whose primary objective was to manipulate the news environment or voter behavior in the run-up to the election.[7] We believe that this model could be further

[7] To add further support to this explanation, we note that of the 96 false-information domains that closed shortly after the election, 86.5% were registered in 2015 or 2016. This compares with false-information websites that remained live, where only 37.8% were registered in 2015 or 2016.

improved with data on actual manipulators rather than the proxy that we employ.[8]

There are several ways in which future implementations of models based on our approach can be improved with further research. First, a future effort may look at ensemble modeling that includes alternative techniques, such as elastic nets, trees and random forests, and mixture models. Second, future implementations may utilize a broader array of features from registration data. For example, as more data becomes available (especially when the test is performed surrounding a particular event), additional terms may be extracted from domain names in continual updates of the model. Third, as an implementation of this type of model becomes iterative, larger amounts of data would improve the predictive power. Our results are achieved with a subset (relative to full browsing history data on all web domains) of imperfect (relative to the information available to ICANN and governmental agencies) data. Additional advances can be expected with better data. Finally, there is an opportunity to complement our approach with unstructured text analysis techniques (for instance, on the first few articles published on the site or the "about" page) to further bolster the predictive power.

Policy makers must balance the need to restrict false information with the need to inappropriately censure alternative views. Such tensions could be mitigated by using our model as a first level filter to identify suspect domains early and target them with more comprehensive (albeit time consuming and expensive) analyses, such as text analysis. In this way, our analysis can be used as a sorting / escalation step in conjunction with other models and sources of information. Doing so can help to improve decision making in the ongoing fight against false information.

References

Acemoglu, D., A. Ozdaglar, and A. ParandehGheibi (2010). Spread of (mis)information in social networks. *Games and Economic Behavior 70*(2), 194–227.

Allcott, H. and M. Gentzkow (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives 31*(2), 211–236.

[8] U.S. and allied intelligence services have confidential data on several domains that they have identified as being directed by foreign entities intent on manipulating the outcome of the election.

Barnes, J., M. Rosenberg, and E. Wong (2020). China and Russia sow disinformation about how U.S. is handling the virus. *The New York Times* (March 29).

Boczkowski, P. J. (2010). *News at Work: Imitation in an Age of Information Abundance.* University of Chicago Press.

Bradshaw, S. and P. N. Howard (2019). *The global disinformation order: 2019 global inventory of organised social media manipulation.* Oxford, UK: Project on Computational Propaganda.

Candogan, O. and K. Drakopoulos (2020). Optimal signaling of content accuracy: Engagement vs. misinformation. *Operations Research 68*(2), 497–515.

Castelo, S., T. Almeida, A. Elghafari, A. Santos, K. Pham, E. Nakamura, and J. Freire (2019). A topic-agnostic approach for identifying fake news pages. In *Companion Proceedings of The 2019 World Wide Web Conference*, pp. 975–980.

Chen, A. (2015). The agency. *The New York Times* (June 5). https://www.nytimes.com/2015/06/07/magazine/the-agency.html.

Conroy, N. K., V. L. Rubin, and Y. Chen (2015). Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology 52*(1), 1–4.

Coppins, M. (2020). The billion-dollar disinformation campaign to reelect the president. *The Atlantic* (March 2020).

Del Vicario, M., A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H. E. Stanley, and W. Quattrociocchi (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences 113*(3), 554–559.

Friedman, J., T. Hastie, and R. Tibshirani (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software 33*(1), 1.

Grinberg, N., K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer (2019). Fake news on twitter during the 2016 U.S. presidential election. *Science 363*(6425), 374–378.

Gu, L., V. Kropotov, and F. Yarochkin (2017). The fake news machine: How propagandists abuse the internet and manipulate the public. Trend Micro, https://documents.trendmicro.com/assets/white_papers/wp-fake-news-machine-how-propagandists-abuse-the-internet.pdf (last accessed Aug 25, 2020).

Guzman, J. and S. Stern (2015). Where is Silicon Valley? *Science 347*(6222), 606–9.

Luca, M. and G. Zervas (2016). Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Science 62*(12), 3412–3427.

Murphy, H. and J. Reed (2020). Facebook accuses telecoms groups of disinformation tactics: South-east Asian providers said to have used fake accounts to discredit rivals. *Financial Times* (February 12).

Papanastasiou, Y. (2020). Fake news propagation and detection: A sequential model. *Management Science 66*(5), 1826–1846.

Rosenberg, M. and J. Barnes (2020). A bible burning, a Russian news agency and a story too good to check out. *The New York Times* (August 12).

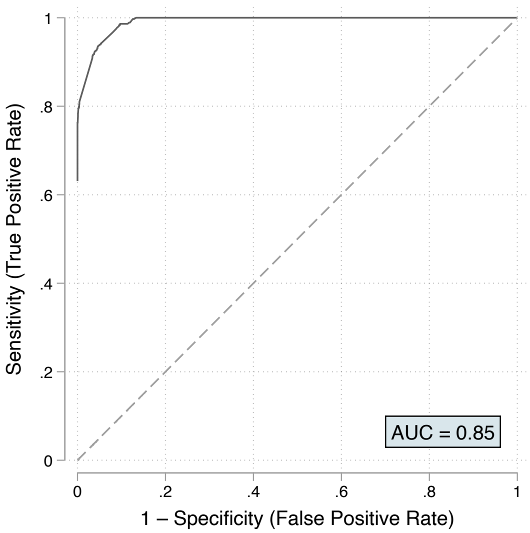
Shu, K., A. Sliva, S. Wang, J. Tang, and H. Liu (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations 19*(1), 22–36.

StatCounter (2016). Desktop, mobile & tablet social media stats United States Of America. https://gs.statcounter.com/social-media-stats/desktop-mobile-tablet/united-states-of-america/#monthly-201601-201612.

U.S. Department of Homeland Security (2019). Combatting targeted disinformation campaigns: A whole-of-society issue. *Public-Private Analytic Exchange Program.*

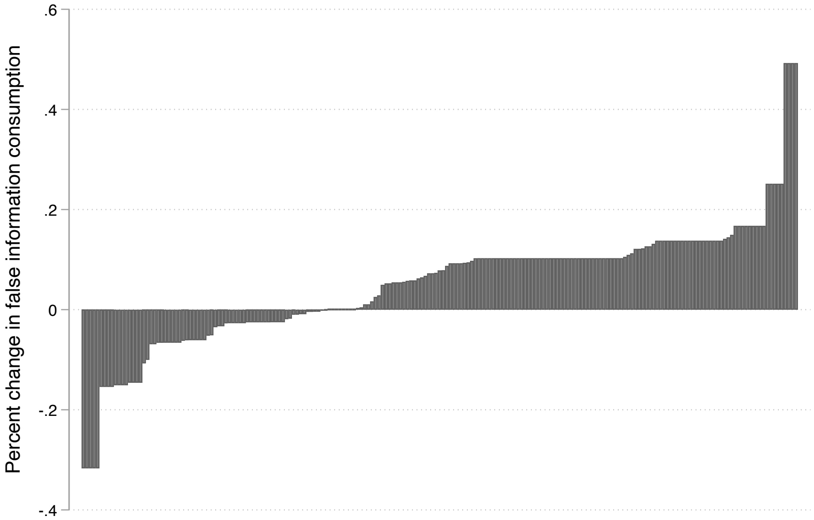
Vosoughi, S., D. Roy, and S. Aral (2018). The spread of true and false news online. *Science 359*, 1146–1151.

Figure 1: Receiver operating characteristic (ROC) curves for false information classifier model



Note: Performance results for our classifier model predicting whether a domain is a false information purveyor among news domains. The ROC curve presents the tradeoff between sensitivity and one minus specificity for cut points, ranging from zero to one. The area under the curve (AUC) is a computation of the area under the ROC curve and represents the performance of the model.

Figure 2: Average change in false-information consumption on the release date of a false information article



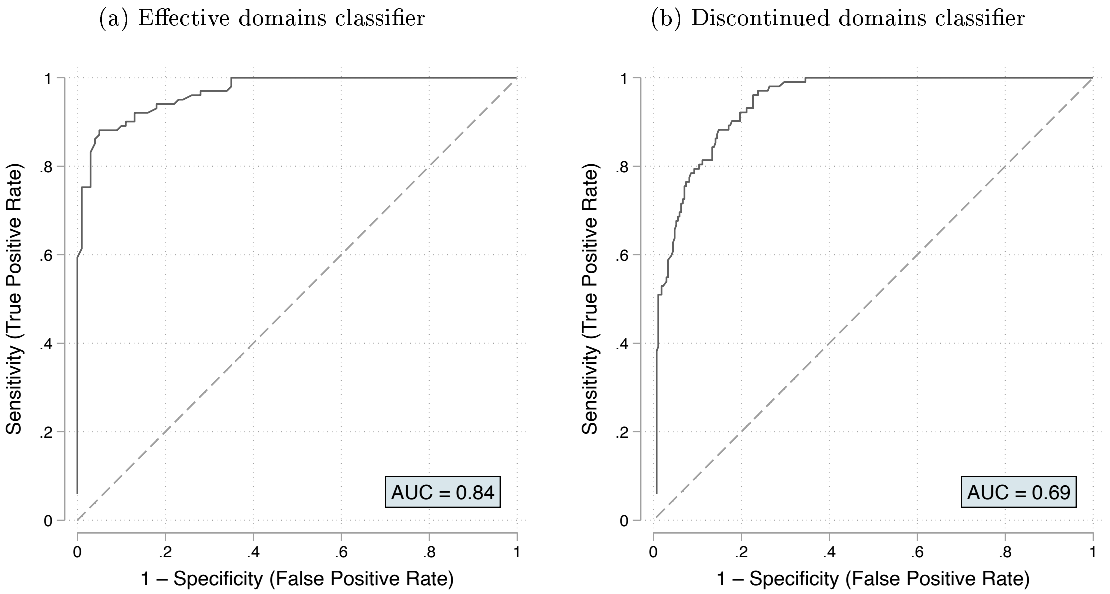
Note: $n = 335$. Each bar represents the average false-information consumption as a percent change over the surrounding five days (from two days prior to two days after) of a false information article on the date it was published.

Table 1: Relationship between false-information production and consumption, by producer's post-election operational status

DV: *False Information Consumption*	(1)	(2)	(3)
Discontinued Production	0.280	0.066	0.072
	(0.135)	(0.108)	(0.103)
Ongoing Production	-0.017	0.012	0.009
	(0.034)	(0.027)	(0.029)
Social Media Visits	0.067	0.015	0.017
	(0.021)	(0.010)	(0.010)
Constant	-1.408	-0.766	-0.781
	(0.196)	(0.122)	(0.148)
Calendar week dummies	Yes	Yes	Yes
Day of week dummies	Yes	Yes	Yes
Chi-Sq Test of discontinued=ongoing	8.94	6.42	6.08
Obs	43,715	21,713	43,715
Users	2,680	1,172	2,680
Log likelihood	-26,964.3	-16,017.6	-21,747.3

Note: This table shows results from Poisson models with robust standard errors (in parentheses) clustered at the user level. Columns 1, 2, and 3 present results from a pooled, fixed effects, and random effects model, respectively. The dependent variable is a count of the number of false information websites visited by a user on the focal day.

Figure 3: Receiver operating characteristic (ROC) curves for false information categorization classifier models



(a) Effective domains classifier

(b) Discontinued domains classifier

Note: (left) Performance results for our classifier model predicting whether a false information domain is efficient in its production of false information.

(right) Performance results for our classifier model predicting whether a domain discontinues operations shortly after the 2016 US election.

The ROC curve presents the tradeoff between sensitivity and one minus specificity for cut points, ranging from zero to one. The area under the curve (AUC) is a computation of the area under the ROC curve and represents the performance of the model.

Online Appendix

A0.1 Sample of General Domains (not false information)

According to DomainTools, 415 million domains were registered from 2006 to 2016. Of those, 70 million were registered in 2016. The following steps were followed to arrive at a sample of general domains for inclusion in our analysis:

1. DomainTools generated a random sample of 75,000 domains whose registration periods roughly approximated the registration periods of the known false information domains. The registration year for the false information domains in our sample are concentrated in 2016, and otherwise dispersed from 2006 through 2015. DomainTools randomly selected 30,000 domains from 2016 registrants and 45,000 domains from 2006-2015 registrants

2. From this set of 75,000 domains, we randomly sampled 4,000 domains for inclusion in our analysis.

3. We used the DomainTools "Who Is History" application programming interface (API) to download registration information on the 4,000 domains, from the first registration event of the domain. Note that DomainTools also maintains information on subsequent renewals, expirations, updates, and third-party acquisitions of each domain.

4. Among the 4,000 domains, complete information was available or could be extracted for 1,861 domains. This represents our final sample of general domains.

A0.2 Data Summary

We summarize the data employed in the regression analysis described in Section 4.3. Individual-level summary statistics—false-information domain visits and social media visits—are provided in Table A1. Summary statistics for day-level variables—publications on domains that were eventually discontinued and on domains that continued to operate—are provided in Table A2.

Table A1: Individual-level summary statistics

	Mean	S.D.	Median	Min	Max
false information consumption	0.60	3.74	0.00	0.00	187.00
social media visits	3.15	22.37	0.00	0.00	1,404.00

Note: $n = 43,715$ user-weekdays for 2,680 users.

Table A2: Daily-level summary statistics

	Mean	S.D.	Median	Min	Max
discontinued false information production	0.23	0.40	0.10	0.00	2.00
ongoing false information production	1.09	1.40	0.80	0.00	6.00

Note: $n = 30$ weekdays. Measures are scaled by a factor of 10.