Identifying Disinformation Websites Using Infrastructure Features

Austin Hounsel
Princeton University

Jordan Holland
Princeton University

Ben Kaiser
Princeton University

Kevin Borgolte
Princeton University

Nick Feamster
University of Chicago

Jonathan Mayer
Princeton University

Abstract

Platforms have struggled to keep pace with the spread of disinformation. Current responses like user reports, manual analysis, and third-party fact checking are slow and difficult to scale, and as a result, disinformation can spread unchecked for some time after being created. Automation is essential for enabling platforms to respond rapidly to disinformation.

In this work, we explore a new direction for automated detection of disinformation websites: infrastructure features. Our hypothesis is that while disinformation websites may be perceptually similar to authentic news websites, there may also be significant non-perceptual differences in the domain registrations, TLS/SSL certificates, and web hosting configurations. Infrastructure features are particularly valuable for detecting disinformation websites because they are available before content goes live and reaches readers, enabling early detection.

We demonstrate the feasibility of our approach on a large corpus of labeled website snapshots. We also present results from a preliminary real-time deployment, successfully discovering disinformation websites while highlighting unexplored challenges for automated disinformation detection.

1 Introduction

In recent years, the Internet has made disinformation cheaper, easier, and more effective than ever before [37]. The same technologies that have democratized online content creation, distribution, and targeting are increasingly being weaponized to mislead and deceive. Russia deployed disinformation to interfere in the 2016 U.S. presidential election [14, 31, 45, 59, 67, 76], and political disinformation campaigns have also struck dozens of other nations [7, 8]. Other disinformation campaigns have been economically motivated, driving page views for advertising revenue, pushing products, or undermining competitors [43].

The major online platforms have not kept pace with the spread of disinformation. Responses to disinformation mostly rely on user reports, manual analysis, and fact checking, which are slow and difficult to scale [5]. Similarly, previous work on automated detection of disinformation has focused on textual and social graph features, which often rely on disinformation already being shared [1, 13, 25, 38, 41, 50, 73, 74, 78]. These responses give disinformation an asymmetric advantage, enabling it to spread and affect perceptions in the hours and days after it is first distributed—a critical period during which disinformation may be most effective [80].

In this paper, we explore the use of infrastructure features to detect disinformation websites. Our hypothesis is that while disinformation websites may be perceptually similar to authentic news websites, there may be significant non-perceptual differences in the domain registrations, TLS/SSL certificates, and web hosting configurations.

We are motivated by prior work in the information security field that demonstrated viable early detection for malware, phishing, and scams using machine learning and a combination of carefully engineered network-level and application-level features [27, 28, 65, 68]. We use similar insights to support discovery of disinformation websites based on infrastructure features. These features are particularly valuable because they are available before disinformation campaigns begin.

We construct features derived from a website's domain, certificate, and hosting characteristics and then apply multi-label classification to categorize the website as disinformation, authentic news, or other (i.e., lacking news content). Our evaluation shows the feasibility of our approach on a large, labeled dataset of website snapshots. We also present results from a preliminary real-time deployment, in which we were able to discover previously unreported disinformation websites. We outline the challenges for automated disinformation detection based on our experience.

2 Definitions and Scope

Definitions of disinformation vary in academic literature and public discourse [37, 64]. Common components include intent to deceive about facts [18, 29, 34, 69], intent to harm [69], and intent to prompt distribution [40].

2.1 Websites as Granularity of Study

We study disinformation at the granularity of website domains, rather than individual articles, claims, advertisements, social media accounts, or social media actions (e.g., posts or shares). Websites are often used as distribution channels for disinformation on social media platforms [2, 3, 22, 24], and there are several benefits to studying disinformation at the website level:

- **Early Warning.** It is possible, in principle, to identify disinformation websites before they begin to publish or distribute content. For example, an automated detection

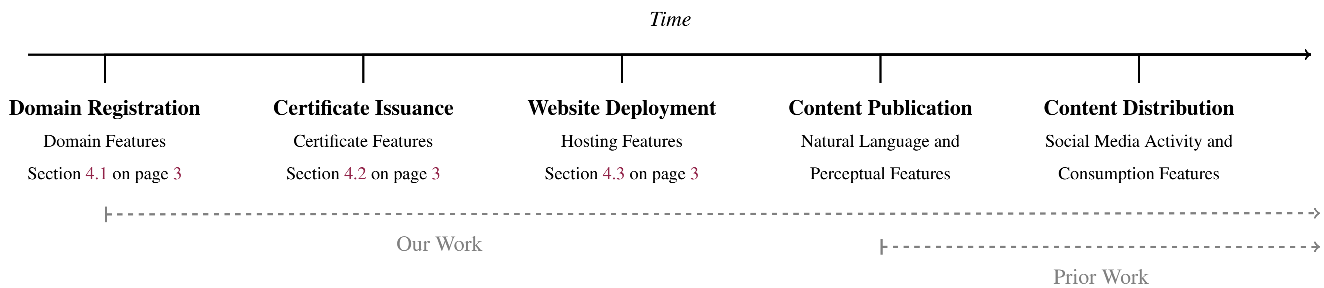


Figure 1: The lifecycle of a disinformation website, from domain registration to content distribution. We focus on features that are available as soon as the domain is registered, with progressively improving automated accuracy as the website deploys.

system might spot a new domain registration that looks like a local newspaper name, but has infrastructure overseas. A human moderator might then investigate and find there is no local newspaper with that name. Analysis of article content or social media activity, by contrast, can only occur much later in the disinformation lifecycle (Figure 1).

- **Longer-Term Value.** Disinformation articles and social media posts have an inherently limited lifespan, owing to the rapid news cycle [80]. Disinformation websites, by contrast, can last for years.
- **Platform Independence.** Identifying disinformation websites is feasible without access to a major online platform's internal account or activity data.
- **Ecosystem Value.** A real-time feed of disinformation websites has value throughout the Internet ecosystem, similar to existing feeds of malware, phishing, and scam domains [20]. Website data is immediately actionable for a diverse range of Internet stakeholders. Further, because websites are often components of multimodal disinformation campaigns, detection at the domain level can provide an investigative thread to untangle the rest of a disinformation campaign, including associated social media accounts and activities.

There are drawbacks and limitations associated with focusing on domains. Some websites feature a mix of authentic and false news, complicating our class definitions (see Section 2.2). We also recognize that websites are just one source of disinformation, and that private communications and native social media content also play substantial roles in exposing users to disinformation and instilling false beliefs.

2.2 Class Definitions

Disinformation We define disinformation websites as websites that appear to be news outlets with content about politics and current events, but that operate in a manner significantly inconsistent with the norms, standards, and ethics of professional journalism. Satire websites fall within our definition of disinformation when the satire is not readily apparent to users. This is an intentional definitional decision, since satire websites can (and often do) mislead users [16, 35], and since disinformation websites are known to sometimes rely on implausible small-print disclaimers that they are satire [47, 48]. Our goal is to identify websites where users might benefit from additional context or other interventions. We decline to use the term "fake news," even though it may be a more apt description of the category of website that we study, because of the term's political connotations and because the utility of our features is generalizable.

We focus on websites related to current events and politics because a significant proportion of the U.S. population has encountered these types of disinformation websites at least once [2, 3, 24], and certain groups of users (such as those over the age of 65) encounter them at high rates [23].

Authentic News We define authentic news websites, including those with a partisan bias, as news websites that adhere to journalistic norms such as attributing authors, maintaining a corrections policy, and avoiding egregious sensationalism.

Non-news We define a third category of non-news websites, which are websites that primarily serve content other than news and do not represent (or claim to represent) news outlets.

3 Website Dataset

We used both current and historical data to construct our dataset. We identified three classes, rather than just two classes of fake and authentic news, both to facilitate feature engineering and because we found that cleaner class separation improved classification performance. We balanced the classes, including about 550 websites for each class.

We first constructed the disinformation class, then constructed the other two sets with equal sizes for balanced training and testing. Class sizes changed slightly over the course of dataset construction, so the final datasets contain 551 disinformation sites, 553 news sites, and 555 non-news sites. We recognize that the non-news website class would predominate in a real-time feed of domain, certificate, or social media events. Our rationale is that without balancing the dataset, the models we develop would minimize error by simply labeling every website as other.

3.1 Disinformation Websites

We began by combining multiple preexisting datasets of disinformation websites that had been manually labeled by experts or published by news outlets, research groups, and professional fact-checking organizations. Specifically, we integrated the corpora from CBS [10], FactCheck.org [17], Snopes [39], Wikipedia [71], PolitiFact [19], and BuzzFeed [61–63]. We also included websites that have been labeled as "disinformation" by OpenSources, a collaborative academic project that manually assigned credibility labels to news-like websites [79]. Finally, we integrated the list of disinformation websites compiled by Allcott et al. for their study on the diffusion of disinformation on Facebook between 2015 and 2018 [3], which they also compiled from lists by fact-checking organizations and academic sources.

We then manually filtered the list of websites, leaving only the websites that satisfied our definition of disinformation (Section 2.2). Our final dataset contains 758 disinformation websites. 575 (76%) of the websites are currently inactive: either unavailable, replaced with a parking page, or repurposed for other kinds of abuse (e.g., spam or malware distribution). This highlights the rapid turnover of disinformation websites in comparison to authentic news websites. Fortunately, we were able to reconstruct domain, certificate, and hosting features for 368 (64%) of these inactive websites through Internet Archive Wayback Machine snapshots, the DomainTools API, and the `crt.sh` Certificate Transparency log database [15, 33, 58]. This resulted in a set of 551 disinformation websites.

3.2 News Websites

We built a corpus of 553 authentic news websites, randomly sampling 275 from Amazon's Alexa Web Information Service (AWIS) [4] and 278 from a directory of websites for local newspapers, TV stations, and magazines [75]. From AWIS, we sampled websites categorized as "news", excluding the 100 most popular websites out of recognition that these websites likely have some distinct properties compared to the long tail of news websites (e.g., high-quality and customized infrastructure). From the local news dataset, we manually filtered to omit websites that did not prominently display news (e.g., TV station websites that served as channel guides).

3.3 Other Websites

We built a set of 555 other websites by sampling from Twitter's Streaming API [66]. We filtered for tweets that contained a URL, extracted the domain name, and then used the Webshrinker classification service [70] to assign labels based on the Interactive Advertising Bureau's standardized website categories [32]. We excluded websites that belonged to the "News" and "Politics" categories.

4 Feature Engineering

We engineered 33 features to distinguish disinformation, authentic news, and other websites (see Table 1). In this section, we describe exemplary features.

4.1 Domain Features

Eighteen features related to a website's domain name, registration, or DNS configuration. Domain names were valuable for distinguishing authentic and disinformation websites from other websites; these classes generally used a domain name with a news-related keyword like "news", "herald", or "chronicle". Domain registrars showed distinct usage patterns between authentic and disinformation websites: two low-cost and consumer-oriented registrars (Namecheap and Enom) were much more common among disinformation websites, while business-oriented registrars like Network Solutions, MarkMonitor, and CSC were more common for authentic news websites. We also found that authentic news websites were much more likely to have old domain registrations with far-off expiration dates, and that disinformation websites were much more likely to use new TLDs like `.news`. We observe that disinformation creators trying to evade detection based on domain features would need to plan in advance, establish business relationships, and bear other costs that may be deterring.

4.2 Certificate Features

Nine features related to a website's TLS/SSL certificate. One key feature was the number of domains that a certificate covers (based on the Subject Alternative Name field). We found that news websites often had more domains in their certificates than disinformation websites, because parent news organizations used one certificate to cover their subsidiaries. We also found, though, that some disinformation websites had a large count of domains in their certificates attributable to low-cost hosting providers that deployed shared certificates.

4.3 Hosting Features

Six features pertained to a website's hosting infrastructure. Using BGP routing tables, we identified the autonomous system hosting each website's IP address and found that mass-market hosting providers like GoDaddy and Namecheap were much more common among disinformation websites, while premium, business-oriented hosting providers like Incapsula were more common among authentic news websites. We also geolocated IP addresses using the MaxMind GeoLite2 database and found that, contrary to our expectations, geolocation was not a valuable feature because most websites in all three classes were U.S.-based.

The presence of web trackers and low-quality advertisements may also be useful features, although we did not find that a feature based on websites' Google Analytics IDs improved model accuracy. We leave the evaluation of other trackers to future work.

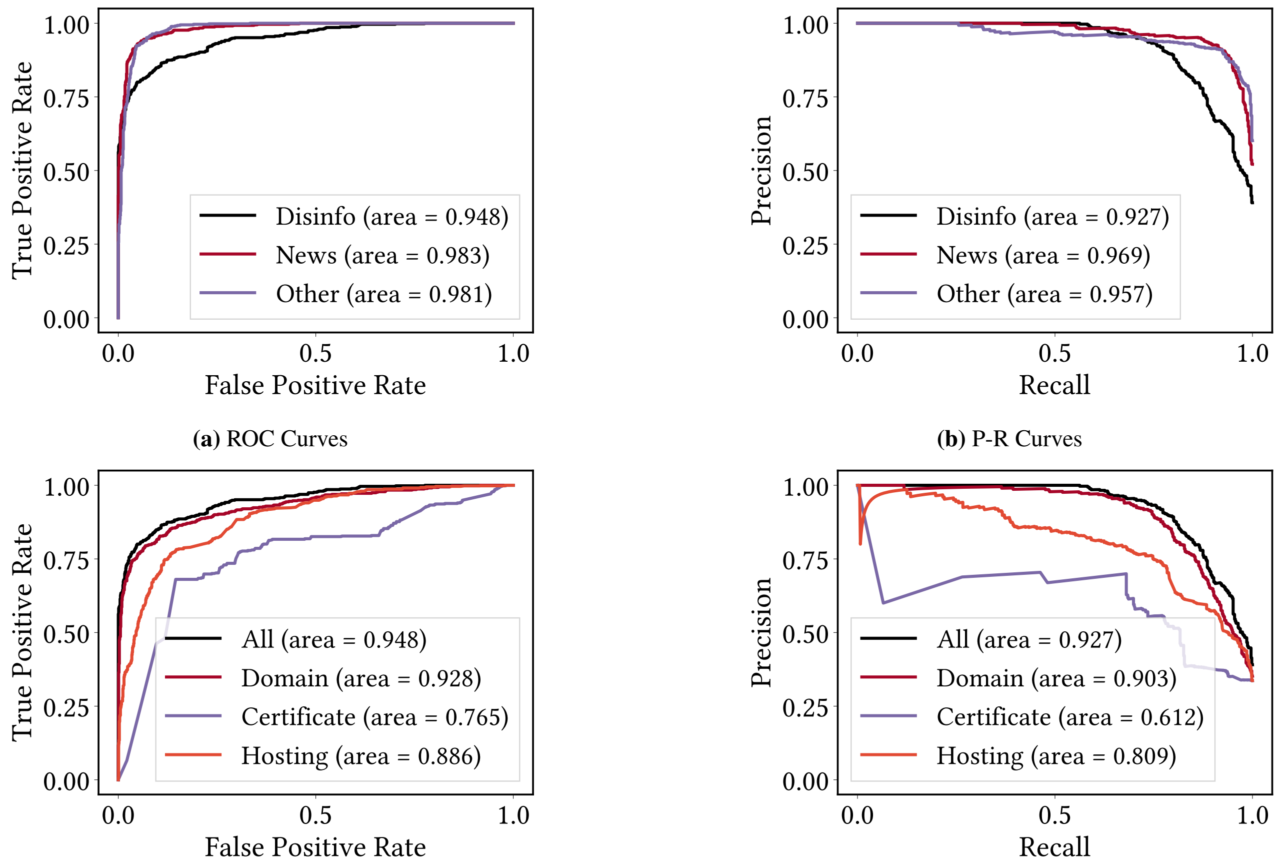
Name	Category	Description	Rank	Data Type
News Keyword(s) in Domain	Domain	The domain name contains one or more keywords that imply it serves news (e.g., “herald,” “tribune,” or “chronicle”).	1	Boolean
Domain Name Length	Domain	The number of characters in the domain name.	3	Numeric
“News” in Domain	Domain	The domain name contains the specific keyword “news.”	8	Boolean
WHOIS Privacy	Domain	The domain registrant is using a WHOIS proxy service or registrar privacy option.	9	Boolean
Registrar Name	Domain	The organization with whom the domain was registered.	11	Categorical
Nameserver SLD	Domain	The second-level domain of the nameserver.	14	Categorical
Nameserver AS	Domain	The autonomous system of the nameserver’s IP address.	16	Categorical
Registrant Organization	Domain	The organization of the registrant.	17	Categorical
Registrant Country	Domain	The country of the registrant.	19	Categorical
Time Since Domain Registration	Domain	The time elapsed since the domain was originally registered.	21	Numeric
Domain Lifespan	Domain	The time period between the domain’s initial registration and expiration dates.	22	Numeric
Time to Domain Expiration	Domain	The time until the domain’s registration expires.	23	Numeric
Time Since Domain Update	Domain	The time since the domain’s configuration was updated.	25	Numeric
Nameserver Country	Domain	The country where the nameserver is located, using IP geolocation.	27	Categorical
Novelty TLD	Domain	The TLD is novelty (e.g., `.news`, `.xyz`, or `.club`).	29	Boolean
Digit in Domain	Domain	The domain name contains numeric characters.	30	Boolean
Hyphen in Domain	Domain	The domain name contains a hyphen.	31	Boolean
Domain Resolves	Domain	The domain name resolves to an IP address.	32	Boolean
SAN Count	Certificate	The number of domains in the Subject Alternate Name extension field.	2	Numeric
SAN Contains Wildcard	Certificate	The Subject Alternate Name extension field contains a wildcard entry for a domain.	7	Boolean
Expired Certificate	Certificate	The certificate is expired.	10	Boolean
Certificate Available	Certificate	A certificate is configured at the domain (i.e., a certificate is provided during a TLS handshake on the HTTPS port).	12	Boolean
Self-signed Certificate	Certificate	The certificate is signed by the domain owner, not a CA.	13	Boolean
Domain-validated Certificate	Certificate	The domain owner has obtained a certificate from a CA through domain validation rather than organization validation or extended validation.	18	Boolean
Certificate Issuer Name	Certificate	The organization or individual who issued the certificate.	24	Categorical
Certificate Issuer Country	Certificate	The country where the certificate was issued.	26	Categorical
Certificate Lifetime	Certificate	The certificate’s period of validity.	28	Numeric
WordPress Plugins	Hosting	WordPress plugins used by the website.	4	Categorical
Website AS	Hosting	The autonomous system of the website’s IP address.	5	Categorical
WordPress CMS	Hosting	The website uses WordPress as its content management system.	6	Boolean
WordPress Theme	Hosting	The WordPress theme used by the website.	15	Categorical
Website Country	Hosting	The country where the website is located, using IP geolocation.	20	Categorical
Website Available	Hosting	A website is hosted at the domain (i.e., content is returned in response to an HTTP request for the base URL, following redirects).	33	Boolean

Table 1: Domain, certificate, and hosting features that our model uses to classify a website as authentic news, disinformation, or other. Features are ranked by Gini importance in our random forest model.

5 Model

We selected a multi-class random forest model because we expected that feature interactions would contribute to performance and that model interpretability would be important for evaluation and plausible deployment. We conducted a randomized hyperparameter search over a wide range of values, then selected values that achieved the best average accuracy based on 250 iterations with five-fold cross-validation.

Our model was able to distinguish the three classes and classify websites accurately; the mean ROC AUCs for authentic news websites, disinformation websites, and other websites were 0.98, 0.95, and 0.98 respectively, and the mean



(a) ROC Curves

(b) P-R Curves

(c) ROC Curves

(d) P-R Curves

Figure 2: Model performance on the curated dataset. Figures 2a and 2b show performance on all three classes using all feature types. Figures 2c and 2d show performance on the disinformation class using subsets of features.

precision-recall AUCs were 0.97, 0.93, and 0.96 respectively. The performance of our model surpasses the prior work [6] and is comparable to concurrent work [12]. We present ROC and precision-recall figures in Figure 2.

We evaluated the importance of the domain, certificate, and hosting feature categories by training and testing a standalone model for each category. Figure 2 presents the comparative performance of these models. We found that domain features predominantly drove classification performance. This result is promising, because domain features are available very early in a disinformation website's lifecycle; domain registrars (among other Internet stakeholders) could intervene or begin heightened monitoring when a suspicious domain appears. We found that hosting features accomplished moderate performance, while certificate features contributed little to classification performance.

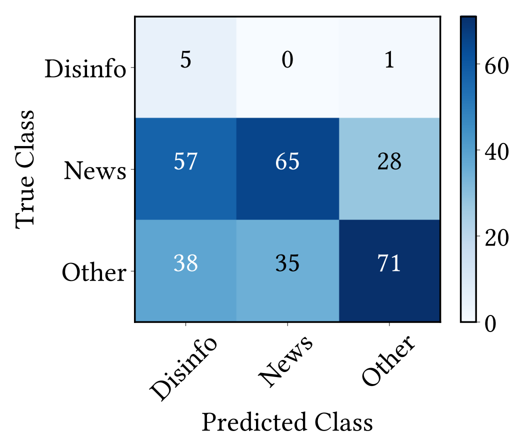


Figure 3: Performance on 100 manually labeled websites from each class, sampled from real-time pilot data.

6 Pilot Real-Time Deployment

We conducted a pilot real-time deployment of our classifier to understand how well its performance generalizes. Our implementation used a commodity server to ingest new domains from DomainTools [15] (which tracks domain registrations), CertStream [9] (which tracks new TLS/SSL certificates), the Twitter Streaming API [66], and the Reddit API [55], then another commodity server to collect infrastructure data, generate features, and output a classification. We ran the system for 5 days, classifying 1,326,151 websites. To simulate our classifier on a platform that has access to all features, we randomly sampled 300 websites that appeared on Twitter—100 from each detected class—and manually labeled to evaluate performance. We present a confusion matrix in Figure 3.

Our model's precision on the disinformation class (0.05) was sufficient for plausible deployment. We were able to rapidly discard false positives and, just in our small-scale pilot deployment, discovered two disinformation websites that had not been previously reported in any public venue.

We cautiously note that model performance radically degraded in comparison to our prior evaluation, similar to results in concurrent work [12]. We attribute this difference to three potential causes. There is a massive class imbalance inherent in real-world classification—the overwhelming majority of websites do not relate to news, and most news links shared on social media are for authentic news websites rather than disinformation websites. As a result, even with good performance, false positives may dominate true positives. Also, as training data gets older, the features for current websites may change. This may create a gap between the features observed in the training data and current disinformation websites.

It is also possible that our model is picking up on artifacts in our training dataset to classify websites. For example, $\sim$34% of the disinformation websites in our training data are active, whereas all of the news websites are active. If there are differences in the features of inactive websites that we reconstructed features for and active websites, then our classifier may be using the wrong signal to distinguish disinformation websites from news websites. We leave an evaluation of our classifier that is only trained on active websites to future work.

7 Evasion Resistance

Disinformation website operators will be motivated to evade detection. In other areas of online abuse, such as spam, phishing, and malware, online platforms are constantly developing new defensive measures to keep up with advances in adversary capabilities. We expect that disinformation will follow a similar cat-and-mouse pattern of defense and evasion.

Our model uses features that provide a degree of asymmetric advantage in identifying disinformation websites, since a website that seeks to evade detection must make changes to its infrastructure. Some features will be relatively easy to evade; for example, a website can easily change a WordPress theme or renew an expired TLS certificate. Fortunately, many of the most important features that our model relies on are difficult or costly to evade.

As an example, consider one of the most predictive features: the lifespan of a website's domain. Evading that feature requires either significant advance planning or purchasing an established domain. Evading certain other features incurs monetary costs, like purchasing a certificate from a reputable issuer, registering a domain for a longer time, switching to a more expensive non-novelty TLD, or migrating to a more trustworthy hosting provider. Evading other features incurs technical costs: obtaining and installing a correctly configured, reputably issued TLS certificate, for instance, imposes some operational cost. Finally, evading many of our model's features might reduce the effectiveness of the disinformation campaign. For example, a top ranked feature is whether a domain contains news keywords. Removing those keywords from the domain name could diminish the credibility of the website and lead to less exposure on social media.

8 Related Work

Our approach is inspired by the information security literature on detecting malicious websites. Prior work has demonstrated the value of infrastructure features for identifying spammers [28, 52], botnets [21, 53], and scams [26, 27, 36].

Efforts at classifying disinformation have predominantly used natural language features (e.g., [1, 11, 13, 30, 38, 42, 44, 46, 49–51, 54, 56, 57, 60, 72, 77]). Several projects have relied on social graph features (e.g., [25, 41, 73, 74, 78]).

Baly et al. predicted news website factuality by examining the domain name, article text, associated Wikipedia and Twitter pages, and web traffic statistics for a combined accuracy of about 0.5 [6]. Chen and Freire developed a system for discovering political disinformation websites on Twitter that uses article text and markup [12]. The authors report a mean ROC AUC of 0.97 on historical data and significantly lower performance in a trial real-time deployment.

In comparison to the disinformation detection literature, we contribute a new set of infrastructure features that do not rely on content or distribution. We demonstrate website classification with state-of-the-art performance and characterize how performance degrades in a real-world deployment setting.

9 Conclusion

Our work demonstrates a promising new set of features for detecting disinformation websites and a real-time classification system could feasibly be deployed by online platforms. Future work on disinformation detection should examine how infrastructure features interact with natural language and social sharing features. Our work also highlights the significant and unexplored practical challenges of real-world deployment for disinformation detection systems. As a research community, we have an opportunity to support the free and open speech environment online by enabling responses to disinformation. But we have to move beyond detection methods that are only strong on paper and begin addressing the hard problems associated with real-world deployment.

Acknowledgments

We thank our paper shepherd Jed Crandall and our anonymous reviewers for their feedback. We also thank DomainTools for providing access to their services for our research. This work was funded in part by National Science Foundation Award TWC-1953513.

References

[1] S. Afroz, M. Brennan, and R. Greenstadt. Detecting Hoaxes, Frauds, and Deception in Writing Style Online. In *Proceedings of the 33rd IEEE Symposium on Security & Privacy (S&P)*, 33rd edition, May 2012. DOI: 10.1109/SP.2012.34.

[2] H. Allcott and M. Gentzkow. Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives (JEP)*, 31(2), 2017. DOI: 10.1257/jep.31.2.211.

[3] H. Allcott, M. Gentzkow, and C. Yu. Trends in the diffusion of misinformation on social media. *Research and Politics*, 6, 2, 2019. DOI: 10.1177/2053168019848554.

[4] Amazon. Alexa Web Information Service. 2020. URL: https://docs.aws.amazon.com/AlexaWebInfoService/latest/index.html.

[5] M. Ananny. The partnership press: Lessons for platform-publisher collaborations as Facebook and news outlets team to fight misinformation. Technical report, The Tow Center for Digital Journalism, Apr. 4, 2018. URL: https://www.cjr.org/tow_center_reports/partnership-press-facebook-news-outlets-team-fight-misinformation.php.

[6] R. Baly, G. Karadzhov, D. Alexandrov, J. Glass, and P. Nakov. Predicting Factuality of Reporting and Bias of News Media Sources. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2018th edition, Oct. 2018. DOI: 10.18653/v1/D18-1389.

[7] S. Bradshaw and P. N. Howard. Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation. Technical report, University of Oxford, July 20, 2018. URL: http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/07/ct2018.pdf.

[8] S. Bradshaw and P. N. Howard. The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation. Technical report, University of Oxford, Sept. 26, 2018. URL: https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf.

[9] Cali Dog Security. CertStream. 2017. URL: https://certstream.calidog.io/.

[10] CBS News. Don't get fooled by these fake news sites. Mar. 2017. URL: https://www.cbsnews.com/pictures/dont-get-fooled-by-these-fake-news-sites/.

[11] Y. Chen, V. L. Rubin, and N. Conroy. Towards News Verification: Deception Detection Methods for News Discourse. In *Proceedings of the 48th European Conference on Principles of Knowledge Discovery and Data Mining (PKDD)*, 48th edition, Jan. 2015. URL: https://ir.lib.uwo.ca/fimspres/46/.

[12] Z. Chen and J. Freire. Proactive discovery of fake news domains from real-time social media feeds. In *Companion Proceedings of the The Web Conference (WWW) 2020*, 29th edition, Apr. 2020. DOI: 10.1145/3366424.3385772.

[13] N. J. Conroy, V. L. Rubin, and Y. Chen. Automatic deception detection: Methods for finding fake news. In *Proceedings of the 78th Association for Information Science and Technology Annual Meeting*, 78th edition, Nov. 2015. DOI: 10.1002/pra2.2015.145052010082.

[14] R. DiResta, K. Shaffer, B. Ruppel, D. Sullivan, R. Matney, R. Fox, J. Albright, and B. Johnson. The Tactics & Tropes of the Internet Research Agency. Technical report, New Knowledge, Columbia University, and Canfield Research, LLC, Dec. 17, 2018. URL: https://cdn2.hubspot.net/hubfs/4326998/ira-report-rebrand_FinalJ14.pdf.

[15] DomainTools. API Overview. 2020. URL: https://www.domaintools.com/resources/api-documentation/brand-monitor/.

[16] D. Evon. Did a Georgia Lawmaker Claim a Chick-fil-A Employee Told Her to Go Back to Her Country? July 24, 2019. URL: https://www.snopes.com/fact-check/georgia-lawmaker-go-back-claim/. Snopes.

[17] FactCheck.org. Misinformation Directory. Nov. 12, 2018. URL: https://www.factcheck.org/2017/07/websites-post-fake-satirical-stories/.

[18] D. Fallis. A Functional Analysis of Disinformation. In *Proceedings of the 9th iConference*, Mar. 2014. DOI: 10.9776/14278.

[19] J. Gillin. PundictFact – PolitiFact's guide to fake news websites and what they peddle. Apr. 20, 2017. URL: https://www.politifact.com/punditfact/article/2017/apr/20/politifacts-guide-fake-news-websites-and-what-they/. PolitiFact.

[20] Google LLC. Google Safe Browsing. 2020. URL: https://safebrowsing.google.com/.

[21] G. Gu, R. Perdisci, J. Zhang, and W. Lee. BotMiner: Clustering Analysis of Network Traffic for Protocol- and Structure-Independent Botnet Detection. In *Proceedings of the 17th USENIX Security Symposium (USENIX Security)*, 17th edition, Aug. 2008. URL: https://www.usenix.org/legacy/event/sec08/tech/full_papers/gu/gu.pdf.

[22] A. Guess, B. Lyons, J. N. Montgomery, B. Nyhan, and J. Reifler. Fake news, Facebook ads, and misperceptions: Assessing information quality in the 2018 U.S. midterm election campaign. Technical report, Princeton University, University of Exeter, Washington University at St. Louis, and University of Michigan, 2018. URL: http://www-personal.umich.edu/~bnyhan/fake-news-2018.pdf.

[23] A. Guess, J. Nagler, and J. Tucker. Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, 5(1), 2019. DOI: 10.1126/sciadv.aau4586.

[24] A. Guess, B. Nyhan, and J. Reifler. Selective Exposure to Misinformation: Evidence from the consumption of fake news during the 2016 U.S. presidential campaign. Technical report, Princeton University, Dartmouth College, and University of Exeter, Jan. 9, 2018. URL: http://www-personal.umich.edu/~bnyhan/fake-news-2016.pdf.

[25] A. Gupta, H. Lamba, P. Kumaraguru, and A. Joshi. Faking Sandy: Characterizing and Identifying Fake Images on Twitter during Hurricane Sandy. In *Proceedings of the 22nd World Wide Web Conference (WWW)*, 22nd edition, May 2013. DOI: 10.1145/2487788.2488033.

[26] S. Hao, N. Feamster, and R. Pandrangi. An Internet-Wide View into DNS Lookup Patterns. Technical report, Georgia Institute of Technology and VeriSign Corporation, 2010. URL: http://www.utdallas.edu/~shao/papers/hao_whitepaper10.pdf.

[27] S. Hao, A. Kantchelian, B. Miller, V. Paxson, and N. Feamster. PREDATOR: Proactive Recognition and Elimination of Domain Abuse at Time-of-Registration. In *Proceedings of the 23rd ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 23rd edition, Oct. 2016. DOI: 10.1145/2976749.2978317.

[28] S. Hao, N. A. Syed, N. Feamster, A. G. Gray, and S. Krasser. Detecting Spammers with SNARE: Spatio-temporal Network-level Automatic Reputation Engine. In *Proceedings of the 18th USENIX Security Symposium (USENIX Security)*, 18th edition, Aug. 2009. URL: https://www.usenix.org/legacy/event/sec09/tech/full_papers/hao.pdf.

[29] P. Hernon. Disinformation and misinformation through the internet: Findings of an exploratory study. *Government Information Quarterly*, 12, 2, 1995. DOI: 10.1016/0740-624X(95)90052-7.

[30] B. D. Horne and S. Adali. This Just In: Fake News Packs A Lot In Title, Uses Simpler, Repetitive Content in Text Body, More Similar To Satire Than Real News. In *Proceedings of the 11th International Conference on Weblogs and Social Media (ICWSM)*, 11th edition, May 2017. URL: https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15772.

[31] P. N. Howard, B. Ganesh, D. Liotsiou, J. Kelly, and C. Francois. The IRA, Social Media, and Political Polarization in the United States, 2012-2018. Technical report, University of Oxford and Graphika, Dec. 17, 2018. URL: https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/12/IRA-Report-2018.pdf.

[32] Interactive Advertising Bureau. IAB Tech Lab Context Taxonomy. 2020. URL: https://www.iab.com/guidelines/iab-quality-assurance-guidelines-qag-taxonomy/.

[33] Internet Archive. Internet Archive Wayback Machine. 2020. URL: https://archive.org/web/.

[34] C. Jack. Lexicon of Lies: Terms for Problematic Information. Technical report, Data & Society Research Institute, Aug. 9, 2017. URL: https://datasociety.net/pubs/oh/DataAndSociety_LexiconofLies.pdf.

[35] A. Kasprak. Did Colin Kaepernick Lobby to Remove the National Anthem from Football? July 24, 2019. URL: https://www.snopes.com/fact-check/kaepernick-remove-national-anthem/. Snopes.

[36] M. Konte, N. Feamster, and J. Jung. Dynamics of Online Scam Hosting Infrastructure. In *Proceedings of the 10th Passive and Active Measurement (PAM)*, volume 5448 of *Lecture Notes in Computer Science (LNCS)*, 10th edition, Apr. 2009. DOI: 10.1007/978-3-642-00975-4_22.

[37] S. Kumar and N. Shah. False Information on Web and Social Media: A Survey, Apr. 23, 2018. arXiv: 1804.08559 [cs.SI].

[38] S. Kumar, R. West, and J. Leskovec. Disinformation on the Web: Impact, Characteristics, and Detection of Wikipedia Hoaxes. In *Proceedings of the 25th World Wide Web Conference (WWW)*, 25th edition, Apr. 2016. DOI: 10.1145/2872427.2883085.

[39] K. LaCapria. Snopes' Field Guide to Fake News Sites and Hoax Purveyors. Jan. 14, 2016. URL: https://www.snopes.com/news/2016/01/14/fake-news-sites/. Snopes.

[40] D. M. J. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, M. Schudson, S. A. Sloman, C. R. Sunstein, E. A. Thorson, D. J. Watts, and J. L. Zittrain. The science of fake news. *Science*, 359(6380), Mar. 9, 2018. DOI: 10.1126/science.aao2998.

[41] J. Ma, W. Gao, and K.-F. Wong. Detect Rumors in Microblog Posts Using Propagation Structure via Kernel Learning. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL)*, 55th edition, Aug. 2017. DOI: 10.18653/v1/P17-1066. (Long paper).

[42] D. M. Markowitz and J. T. Hancock. Linguistic Traces of a Scientific Fraud: The Case of Diederik Stapel. *PLoS One*, 9, 8, Aug. 2014. DOI: 10.1371/journal.pone.0105937.

[43] A. Marwick and R. Lewis. Media Manipulation and Disinformation Online. Technical report, Data & Society Research Institute, May 15, 2017. URL: https://datasociety.net/pubs/oh/DataAndSociety_MediaManipulationAndDisinformationOnline.pdf.

[44] R. Mihalcea and C. Strapparava. The Lie Detector: Explorations in the Automatic Recognition of Deceptive Language. In *Proceedings of the 4th International Joint Conference on Natural Language Processing (IJCNLP)*, 4th edition, Aug. 2009. DOI: 10.3115/1667583.1667679. (Short paper).

[45] R. S. Mueller III. Report On The Investigation Into Russian Interference In The 2016 Presidential Election. Mar. 2019. URL: https://www.justice.gov/storage/report.pdf. United States Department of Justice.

[46] N. O'Brien, S. Latessa, G. Evangelopoulos, and X. Boix. The Language of Fake News: Opening the Black-Box of Deep Learning Based Detectors. In *Proceedings of the 2018 Workshop on AI for Social Good*, Nov. 2018. DOI: 1721.1/120056. Co-located with the Conference on Neural Information Processing Systems (NIPS).

[47] B. Palma. Did Minnesota Schools Make Arabic Classes Mandatory? July 19, 2019. URL: https://www.snopes.com/fact-check/minnesota-schools-arabic/. Snopes.

[48] B. Palma. Fake News Purveyors Apologize for 'Satire' Story Calling Fallen Soldier a Deserter. Oct. 30, 2017. URL: https://www.snopes.com/news/2017/10/30/fake-news-purveyors-apologize-satire-story-calling-fallen-soldier-deserter/. Snopes.

[49] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea. Automatic Detection of Fake News, Aug. 23, 2017. arXiv: 1708.07104.

[50] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, and B. Stein. A Stylometric Inquiry into Hyperpartisan and Fake News. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL)*, 56th edition, July 2016. URL: https://www.aclweb.org/anthology/P18-1022.pdf. (Long paper).

[51] V. Qazvinian, E. Rosengren, D. R. Radev, and Q. Mei. Rumor Has It: Identifying Misinformation in Microblogs. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2011th edition, July 2011. URL: http://aclweb.org/anthology/D11-1147.

[52] A. Ramachandran and N. Feamster. Understanding the Network-Level Behavior of Spammers. In *Proceedings of the 2006 ACM SIGCOMM Conference (SIGCOMM)*, Aug. 2006. DOI: 10.1145/1159913.1159947.

[53] A. Ramachandran, N. Feamster, and D. Dagon. Revealing Botnet Membership Using DNSBL Counter-Intelligence. In *Proceedings of the 2nd Workshop on Steps to Reducing Unwanted Traffic on the Internet (SRUTI)*, 2nd edition, July 2006. URL: https://www.usenix.org/legacy/event/sruti06/tech/full_papers/ramachandran/ramachandran.pdf.

[54] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi. Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2017th edition, Sept. 2017. DOI: 10.18653/v1/D17-1317.

[55] Reddit. reddit.com: api documentation. 2020. URL: https://www.reddit.com/dev/api/.

[56] V. L. Rubin, Y. Chen, and N. J. Conroy. Deception detection for news: Three types of fakes. In *Proceedings of the 78th Association for Information Science and Technology Annual Meeting*, 78th edition, Nov. 2015. DOI: 10.1002/pra2.2015.145052010083.

[57] V. L. Rubin and T. Lukoianova. Truth and deception at the rhetorical structure level. *Journal of the Association for Information Science and Technology*, 66, 5, June 5, 2014. DOI: 10.1002/asi.23216.

[58] Sectigo Limited. crt.sh | Certificate Search. 2020. URL: https://crt.sh/.

[59] Select Committee on Intelligence. Report on Russian Active Measures Campaigns and Interference in the 2016 U.S. Election, Volume 2: Russia's Use of Social Media. Oct. 2019. URL: https://www.intelligence.senate.gov/sites/default/files/documents/Report_Volume2.pdf. United States Senate.

[60] K. Shu, D. Mahudeswaran, and H. Liu. FakeNewsTracker: a tool for fake news collection, detection, and visualization. In *Proceedings of the 2018 International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction and Behavior Representation in Modeling and Simulation*, 11th edition, July 2018. DOI: 10.1007/s10588-018-09280-3.

[61] C. Silverman. Here Are 50 Of The Biggest Fake News Hits On Facebook From 2016. Dec. 30, 2016. URL: https://www.buzzfeednews.com/article/craigsilverman/top-fake-news-of-2016. BuzzFeed News.

[62] C. Silverman, J. Lytvynenko, and S. Pham. These Are 50 Of The Biggest Fake News Hits On Facebook In 2017. Dec. 28, 2017. URL: https://www.buzzfeednews.com/article/craigsilverman/these-are-50-of-the-biggest-fake-news-hits-on-facebook-in. BuzzFeed News.

[63] C. Silverman, J. Singer-Vine, and L. T. Vo. In Spite Of The Crackdown, Fake News Publishers Are Still Earning Money From Major Ad Networks. Apr. 4, 2017. URL: https://www.buzzfeednews.com/article/craigsilverman/fake-news-real-ads. BuzzFeed News.

[64] E. C. Tandoc Jr., Z. W. Lim, and R. Ling. Defining 'Fake News'. *Digital Journalism*, 6, 2, Feb. 2018. DOI: 10.1080/21670811.2017.1360143.

[65] K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song. Design and Evaluation of a Real-Time URL Spam Filtering Service. In *Proceedings of the 32nd IEEE Symposium on Security & Privacy (S&P)*, 32nd edition, May 2011. DOI: 10.1109/SP.2011.25.

[66] Twitter. Filter realtime Tweets. Oct. 14, 2019. URL: https://developer.twitter.com/en/docs/tweets/filter-realtime/overview.

[67] United States of America, Department of Justice. Indictment, United States v. Internet Research Agency LLC, Feb. 16, 2018. URL: https://www.justice.gov/file/1035477/download. No. 1:18-cr-00032-DLF (D.D.C. Feb. 16, 2018).

[68] G. Wang, J. W. Stokes, C. Herley, and D. Felstead. Detecting malicious landing pages in Malware Distribution Networks. In *Proceedings of the 43rd IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, June 2013. DOI: 10.1109/DSN.2013.6575316.

[69] C. Wardle and H. Derakhshan. Information Disorder: Toward an interdisciplinary framework for research and policy making. Technical report, Council of Europe, Sept. 27, 2017. URL: https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c.

[70] Webshrinker. APIs - Webshrinker. 2020. URL: https://www.webshrinker.com/apis/.

[71] Wikipedia. List of fake news websites. Jan. 27, 2019. URL: https://en.wikipedia.org/w/index.php?title=List_of_fake_news_websites&oldid=880357562. Page Version ID: 880357562.

[72] G. Wu, D. Greene, B. Smyth, and P. Cunningham. Distortion as a Validation Criterion in the Identification of Suspicious Reviews. In *Proceedings of the First Workshop on Social Media Analytics*, July 10, 2010. URL: http://snap.stanford.edu/soma2010/papers/soma2010_2.pdf.

[73] K. Wu, S. Yang, and K. Q. Zhu. False rumors detection on sina weibo by propagation structures. In *Proceedings of the 31st International Conference on Data Engineering (ICDE)*, 31st edition, Apr. 2015. DOI: 10.1109/ICDE.2015.7113322.

[74] F. Yang, Y. Liu, X. Yu, and M. Yang. Automatic detection of rumor on Sina Weibo. In *Proceedings of the 2012 ACM Workshop on Automated Decision Making for Active Cyber Defense (SafeConfig)*, Aug. 2012. DOI: 10.1145/2350190.2350203.

[75] L. Yin. Local News Dataset. Aug. 14, 2018. URL: https://zenodo.org/record/1345145.

[76] S. Zannettou, T. Caulfield, E. De Cristofaro, M. Sirivianos, G. Stringhini, and J. Blackburn. Disinformation Warfare: Understanding State-Sponsored Trolls on Twitter and Their Influence on the Web. In *Companion Proceedings of the The Web Conference (WWW) 2019*, 28th edition, May 2019. DOI: 10.1145/3308560.3316495.

[77] H. Zhang, Z. Fan, J. Zheng, and Q. Liu. An Improving Deception Detection Method in Computer-Mediated Communication. *Journal of Networks*, 7, 11, Nov. 2012. DOI: 10.4304/jnw.7.11.1811-1816.

[78] Z. Zhao, P. Resnick, and Q. Mei. Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts. In *Proceedings of the 24th World Wide Web Conference (WWW)*, 24th edition, May 2015. DOI: 10.1145/2736277.2741637.

[79] M. Zimdars. OpenSources. Profesionally curated lists of online sources, available free for public use. Apr. 28, 2017. URL: http://www.opensources.co/.

[80] A. Zubiaga, M. Liakata, R. Procter, G. Wong Sak Hoi, and P. Tolmie. Analysing How People Orient to and Spread Rumours in Social Media by Looking at Conversational Threads. *PLoS One*, 11, 3, Mar. 2016. DOI: 10.1371/journal.pone.0150989.

A Appendix

A.1 Domain Proxy Keywords

We first computed the most popular values of the "registrant organization" field from WHOIS records in our training data. Then, we manually extracted keywords from known domain proxy services that were highly ranked. We used the resulting list of keywords as a heuristic to determine if new domains use WHOIS privacy services:

domain	protect	whois
guard	proxy	
privacy	redacted	

A.2 News-Indicating Keywords

We used the following 163 keywords with the DomainTools Brand Monitor API to retrieve newly registered domains that may be news websites. We derived the keywords by manually examining our training datasets of authentic and fake news websites for common terms related to news, media, information, or publishing.

24	365	abc	action
activist	advance	alert	alliance
alternative	america	associate	blast
blog	box	breaking	brief
bulletin	business	buzz	byte
caller	cbs	channel	christian
chronicle	citizen	city	club
cnn	conservative	corner	county
courier	currant	daily	democracy
dig	dispatch	division	edition
editor	election	empire	epoch
evening	examiner	express	extra
fact	feed	file	finesser
flash	focus	fox	free
fresh	gazette	global	guardian
hangout	headline	herald	hq
hub	idea	independent	index
info	inquire	insider	interesting
international	item	journal	leak
learn	ledger	liberal	liberty
live	local	mag	maga
mail	media	metro	movement
nation	nbc	network	now
observer	page	paper	patriot
pioneer	pipe	plug	politic
post	press	progress	proud
publish	radio	react	read
record	region	religion	report
republic	review	rumor	scoop
sentinel	share	sharing	show
spirit	spot	spotlight	standard
star	state	statesman	stories
story	studio	sun	surge
syndicate	telegram	telegraph	television
times	today	top	trend
tribune	truth	tv	twenty-four
twentyfour	uncut	underground	union
update	us	usa	view
viral	vision	washington	watch
weekly	wire	your	zine
zone			