






# TROPIC – Trustworthiness Rating of Online Publishers Through Online Interactions Calculation

Manuel Pratelli<sup>1,2</sup> , Fabio Saracco<sup>1,3,4</sup> , and Marinella Petrocchi<sup>1,2</sup> 

<sup>1</sup> IMT School for Advanced Studies Lucca, Lucca, Italy

<sup>2</sup> CNR-IIT, Pisa, Italy

[manuel.pratelli@imtlucca.it](mailto:manuel.pratelli@imtlucca.it)

<sup>3</sup> ‘Enrico Fermi’ Research Center (CREF), Rome, Italy

<sup>4</sup> CNR-IAC, Sesto Fiorentino, Italy

**Abstract.** Existing methods for assessing the trustworthiness of news publishers face high costs and scalability issues. The tool presented in this paper supports the efforts of specialized organizations by providing a solution that, starting from an online discussion, provides (i) trustworthiness ratings for previously unclassified news publishers and (ii) an interactive platform to guide annotation efforts and improve the robustness of the ratings. The system implements a novel framework for assessing the trustworthiness of online news publishers based on user interactions on social media platforms.

**Keywords:** Online Publisher Trustworthiness · Social Network Analysis

## 1 Introduction

The decline of traditional journalism, coupled with the increasing challenges of maintaining high standards of editorial quality, has raised concerns about the trustworthiness of online news media. Specialized organizations [1, 6–9] provide valuable ratings that assess the credibility of digital news publishers. Although these ratings provide beneficial insights, they are based on the evaluation of resource-intensive and time-consuming criteria, such as the presence of biased and propagandist content in the news articles [10, 11]. As a result, many online publishers remain unlabeled, creating a gap in coverage.

**Proposal:** Based on the work in [13], this demo paper presents TROPIC– Trustworthiness Rating of Online Publishers Through Online Interactions Calculation (available at <https://tropic.iit.cnr.it>), an interactive tool to improve current methods for assessing the trustworthiness of an online publisher and to avoid potential problems of limited coverage. TROPIC outputs include:

- **The automatic classification of the level of trustworthiness of an online news publisher:** TROPIC provides an assessment of the trustworthiness of unclassified online news publishers by analyzing social media interactions between news producers and consumers. The evaluation starts with the analysis of an online social discussion and leverages (i) a *base-knowledge* consisting of a subset of annotated news publishers, (ii) the concept of *News Engagement Communities*, NECs for brevity, which are communities of news articles that received the most attention from discussion participants, introduced in [12, 13] and (iii) the propensity of social users to share low-quality information. This approach provides (i) good accuracy-cost trade-off and (ii) good coverage of *unclassified* and *untrustworthy* news publishers [13].
- **An interactive guide for base-knowledge extension:** The robustness of the classification can be continuously improved by gradually extending the current *base-knowledge*. The intuition is that an extended knowledge allows for more accurate social user profiling, and thus more accurate predictions about the trustworthiness of unannotated news publishers. To guide the expansion of the *base-knowledge*, TROPIC provides the end user with a set of functionalities that suggest the “best” next publisher to annotate. This guides where to focus the extension of the base knowledge, taking best advantage of the prediction system.

**Context:** Manual methods of assessing the trustworthiness of online newspapers are costly, both in terms of the time required to perform the assessment and the cost of finding experienced reviewers. These barriers limit the reach of assessable newspapers. The low coverage also stems from the constant emergence of new, lesser-known online newspapers. TROPIC aims to streamline the rating process and expand the reach of rated newspapers.

**Audience:** Designed for journalists, media professionals, and researchers evaluating the quality of online information, TROPIC provides estimates of the trustworthiness of online publishers. The tool guides the user in the selection of online publishers to manually annotate, so that the subsequent automatic evaluation of news publishers not yet evaluated is effective in terms of prediction accuracy and coverage of the largest number of news publishers evaluated.

## 2 System Overview

**External Data:** Figure 1 shows the tool’s architecture. TROPIC takes as input an online discussion, in the format *edge-list*, i.e., a list of URL - user ID pairs. The URL points to an online news story and is present in social posts, the user ID is the username of the social user who posted or shared that URL on the social platform. The tool is adaptable to data from any social media platform<sup>1</sup>. A second (optional) input for TROPIC is the initial base-knowledge, which is a list of online publishers with associated trustworthiness scores (integers

<sup>1</sup> URL preparation, such as resolving short URLs, is left to the TROPIC user.

from 0 to 100). Since TROPIC is designed to support organizations specializing in evaluating news publishers’ trustworthiness, the initial knowledge base can consist of publishers already assessed by these organizations<sup>2</sup>.

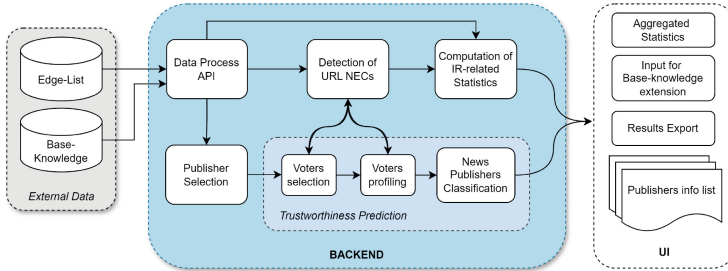


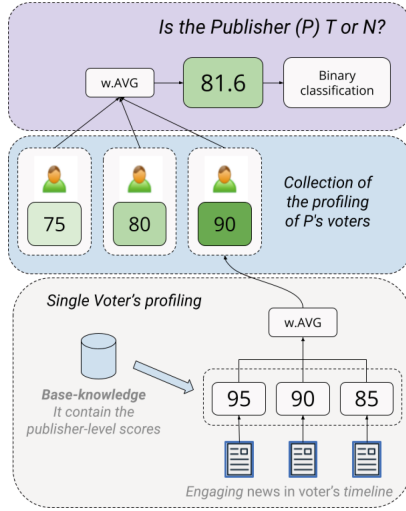
Fig. 1. System Overview

**Backend:** As we can observe in Fig. 1, the TROPIC backend consists of independent software components that work together to provide (i) a predicted trustworthiness score (with a corresponding confidence level) for each publisher not in the base-knowledge, and (ii) metrics to guide users on which publishers to annotate in case they want to manually expand the base-knowledge. The edge list is used both to build NECs communities (which we recall are communities of news URLs that have been very successful among online users in terms of engagement [12, 13]) and to calculate some quantity related to the online discussion we are studying (IR-related statistics). Once we have selected the publishers whose URLs have been shared in the online discussion, we move on to selecting a special group of users, called voters, and profiling them. After this profiling phase, these voters are asked to classify the publishers they interact with.

In our current implementation:

1. We select as voters those users who have shared at least one engaging news (i.e., a news whose URL belongs to one of the NECs).
2. For profiling voters - i.e., estimating each voter’s propensity to share low-quality content - we follow the approach in [13], which consists of a two-step method to downscale the problem. First, instead of considering the complete set of URLs in a voter’s timeline, we only select engaging URLs, specifically those belonging to NECs. Second, instead of evaluating each individual online news article, we evaluate the trustworthiness of the URL by referencing the publisher-level score, if available, from the base knowledge (e.g., source-level scores of 95, 90, and 85, as shown in Fig. 2, bottom). These publisher-level scores are aggregated using a weighted average, resulting in a single final score for each voter (middle part in Fig. 2).

<sup>2</sup> For example, NewsGuard [9] uses the 0–100 scale, the same as TROPIC. In contrast, other organizations, such as Media Bias / Fact Check [8], use descriptive labels and must be converted to numerical values to be used in TROPIC.



**Fig. 2.** Publisher Trustworthiness Classification

Finally, to compute the trustworthiness score for an unclassified publisher, we select all voters who share content from that publisher, profile them, and aggregate their scores into a single (average) number that expresses the trustworthiness score of the considered publisher (Fig. 2, top).

**User Interface:** Figure 3 shows the User Interface (UI). The toolbar allows to upload the edge list (via the “File Upload” button) and, optionally, the *base-knowledge*, and to initiate the computation via the "Process data" button. Calculation results are displayed in three plots and a table. The table lists each publisher’s Name, IR-related statistics, an editable Annotated Score field (this allows the user to manually annotate other domains, if desired, to extend the base knowledge), a Trustworthiness column (showing both binary labels - trustworthy T/ untrustworthy N) and prediction scores, a Confidence column, and a State column (indicating if the rating comes from human annotation or from TROPIC - A stands for annotated, P stands for Predicted). Users can sort the table by IR-related metrics, such as the number of voters or the number of news URLs shared within NECs. This way, manual annotations can be prioritized based on a metric of interest. Above the table, plots provide information on the number of manually annotated publishers and the distribution of their scores (left plot); the number of publishers annotated by human experts (dark color at bottom) and those whose score are calculated by TROPIC (light color at top) (middle plot); and the confidence level of the scores calculated by TROPIC (right plot). This information guides the user to minimize the number of publishers that remain unclassified and to improve the confidence level of the predictions (by adding new manual annotations if necessary). The “Export Actual Knowledge” button allows to export the calculation results in CSV format.

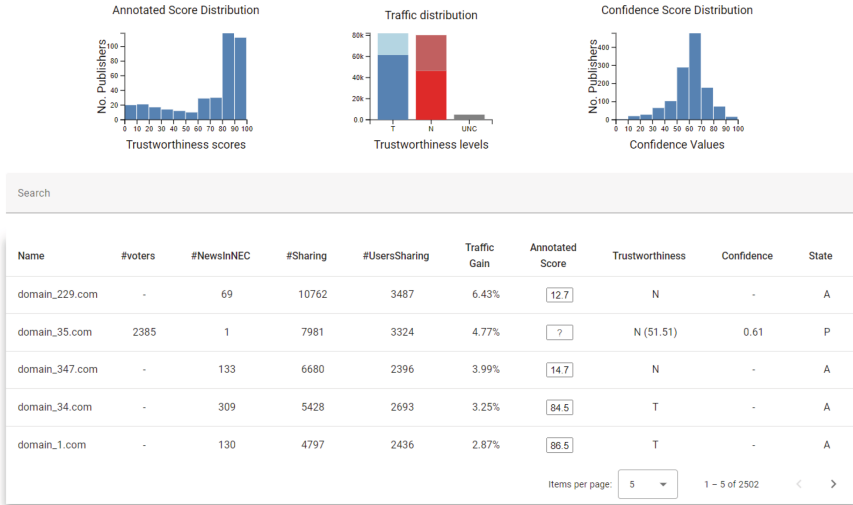


Fig. 3. The User Interface

**Implementation and Demonstration:** The user interface is built with Angular [2], and the backend uses FastAPI [5]. NECs extraction employs the bicm Python library [3]. Both UI and backend are containerized separately with Docker [4]. For practical testing, we provide the tool in a DEMO version. In this configuration, the waiting times for calculating the NECs, which could be several minutes for a very large edge list, are pre-calculated. To utilize this setup, the end user is supplied with a DEMO edge list, which must be selected before pressing “Process Data”. For completeness, a DEMO base-knowledge is also provided, featuring randomly generated trust scores. If end users want to upload their own edge list, the number of edges that can be uploaded in the demo version is limited to 50,000 entries.

### 3 Conclusions

In this demo paper, we presented an interactive web interface designed to streamline the annotation process for assessing the trustworthiness score of an online news publisher. TROPIC predicts trustworthiness scores for publishers that have not yet been manually labeled, improving the coverage and reliability characterization of news domains traffic exchanged in an online discussion.

**Acknowledgments.** This work is partially supported by SERICS (PE00000014) under the NRRP MUR program funded by the EU - #NGEU. FS was partially supported by the project “CODE – Coupling Opinion Dynamics with Epidemics”, funded under NRRP MUR Mission 4 “Education and Research” - Component C2 - Investment 1.1 - Next Generation EU “Fund for National Research Program and Projects of Significant National Interest” PRIN 2022 NRRP, grant code P2022AKRZ9.

## References

1. Ad fontes media. <https://adfontesmedia.com/>, Accessed 23 October 2024
2. Angular. <https://angular.dev/>, Accessed 23 October 2024
3. Bicm. <https://github.com/mat701/BiCM>, Accessed 23 October 2024
4. Docker. <https://www.docker.com/>, Accessed 23 October 2024
5. Fastapi. <https://fastapi.tiangolo.com>, Accessed 23 October 2024
6. The global disinformation index. <https://www.disinformationindex.org/>, Accessed 23 October 2024
7. Iffy index. <https://iffy.news/index/>, Accessed 23 October 2024
8. Mediabias fact check. <https://mediabiasfactcheck.com/>, Accessed 23 October 2024
9. Newsguard. <https://www.newsguardtech.com>, Accessed 23 October 2024
10. Bazmi, P., Asadpour, M., Shakery, A.: Multi-view co-attention network for fake news detection by modeling topic-specific user and news source credibility. *Inform. Process. Manag.* **60**(1), 103146 (2023)
11. Kim, A., Moravec, P., Dennis, A.: Combating fake news on social media with source ratings: the effects of user and expert reputation ratings. *J. Manag. Inf. Syst.* **36**(3), 931–968 (2019). <https://doi.org/10.1080/07421222.2019.1628921>
12. Pratelli, M., Saracco, F., Petrocchi, M.: Entropy-based detection of Twitter echo chambers. *PNAS Nexus* **3**(5), pgae177 (04 2024). <https://doi.org/10.1093/pnasnexus/pgae177>
13. Pratelli, M., Saracco, F., Petrocchi, M.: Unveiling news publishers trustworthiness through social interactions. In: *Proceedings of the 16th ACM Web Science Conference, WEBSCI 2024*, pp. 139–148. Association for Computing Machinery, New York (2024). <https://doi.org/10.1145/3614419.3644015>