TrustLab$^{\text{TM}}$: An Interactive Tool for Evaluating Online Trustworthiness across Diverse Domains

Teng-Chieh Huang, Wenting Song, Brian Kim, and K. Suzanne Barber
The University of Texas at Austin
The Center for Identity
Austin, Texas, USA
e-mail: tengchieh@utexas.edu, wentingsong@utexas.edu, bkim994@utexas.edu, and sbarber@identity.utexas.edu

***Abstract*—TrustLab$^{\text{TM}}$ is an innovative online tool for assessing social media users' trustworthiness with ease and precision. It serves diverse users, from general social media participants to researchers aiming to gauge trust levels in various domains. Unlike many tools, TrustLab$^{\text{TM}}$ focuses on user trustworthiness rather than post content, distinguishing between experts and typical users. Using Trust Filters and user attributes, it assigns trust scores visualized through intuitive charts for clarity. Additionally, TrustLab$^{\text{TM}}$ provides personalized recommendations to help users enhance their online credibility. While its algorithms are domain-independent, this paper demonstrates TrustLab$^{\text{TM}}$'s application in finance, politics, and health, showcasing its role in shaping public discourse, knowledge, and connections. With its user-friendly interface, TrustLab$^{\text{TM}}$ is a significant tool for exploring and understanding online trust in the digital era.**

***Index Terms*—online trustworthiness; individual trust; social media; Trust Filter.**

I. Introduction

Social media has transformed global communication, enabling rapid information sharing and connection. Platforms like X (formerly Twitter, 2006) popularized microblogging, while LinkedIn (2003) fostered professional networking, and Instagram (2010) and Snapchat (2011) emphasized visual sharing. Recently, TikTok (2016) has gained popularity for its short-form videos.

Alongside these benefits, social media has amplified the spread of misinformation, which can significantly impact public opinion on health, politics, and social issues. Misinformation on health topics, for instance, may discourage proper medical actions, while disinformation campaigns can manipulate public perception, influence elections, and create social discord. The rise of generative artificial intelligence (GenAI) further complicates trust by enabling realistic but misleading content, such as deepfake videos [1], [2].

To counter misinformation, platforms like X have implemented tools such as content flagging and partnerships with fact-checkers. Researchers also focus on approaches like rumor detection [3], [4] and bot detection [5]–[8]. This study introduces TrustLab$^{\text{TM}}$, a tool that uses six Trust Filters—authority, experience, expertise, identity, proximity, and reputation—to score user trustworthiness and rank users accordingly. TrustLab$^{\text{TM}}$ uses social media activity and user profiles to enhance trust in online discourse.

The contributions of this study are:

- Development of TrustLab$^{\text{TM}}$, an interactive tool scoring user trustworthiness.
- Application of TrustLab$^{\text{TM}}$ Trust Filters across finance, politics, and health domains.
- Comparison of TrustLab$^{\text{TM}}$ with other trust-evaluation tools, highlighting technology and method distinctions.

The remainder of this paper details the capabilities and algorithms of TrustLab$^{\text{TM}}$ (Section II), compares it with related works (Section III), and discusses future directions (Section IV).

II. Exploring TrustLab$^{\text{TM}}$: Capabilities, Techniques, and Competitive Landscape

TrustLab$^{\text{TM}}$ applies machine learning to score social media posts using Trust Filters based on attributes like average post length, follower count, and post frequency. For instance, in finance and politics, a random forest regressor combined with these attributes showed high accuracy in assessing post trustworthiness [9]. We also found that different attributes are critical across domains: for example, restrained language is key in finance, while post frequency and experience are pivotal in health [10], [11].

Sentiment analysis, particularly on posts with emoticons, has proven effective for identifying trustworthy users, achieving over 80% accuracy [12]. The TrustLab$^{\text{TM}}$ Sentiment Trust Filter, for instance, has helped investors improve returns by following trustworthy sources. In health, Trust Filters were used to detect and forecast disease outbreaks, with epidemiologists tracing posts to monitor potential disease spread [13].

TrustLab$^{\text{TM}}$ has demonstrated positive results across diverse domains, such as finance, politics, and health [9]–[11]. This study focuses on these three areas where misinformation often affects user decision-making, underscoring the importance of high-quality information for better choices.

A. TrustLab$^{\text{TM}}$ Trust Filters: Scoring the Trust

TrustLab$^{\text{TM}}$ scores user trustworthiness through filters like Experience, Reputation, Expertise, Authority, Identity, and Proximity. Each filter assigns scores (0-1) based on specific criteria, such as a user's social links, proximity to an event, or expertise level [11]. Additional attributes, like post length

ISBN: 978-1-68558-236-4

and punctuation use, further refine trust scores [9], [10]. TrustLab$^{\text{TM}}$ has shown high predictive accuracy (exceeding 99.99%) in identifying finance experts and users in financial discussions [10]. The algorithm also achieved accurate predictions in political events, such as the 2016 and 2020 U.S. elections, providing near-instant results at low computational cost.

B. Target Domains Selection

To showcase TrustLab$^{\text{TM}}$'s versatility, we selected finance, health, and politics. These domains are highly influential, have substantial online misinformation, and can benefit from improved information quality. In finance, understanding public sentiment around stock market trends can help anticipate market movements. For politics, social media analysis has provided cost-effective and timely election forecasts [14]. In health, TrustLab$^{\text{TM}}$ has been used to monitor disease spread through trusted social media posts, enabling early response in biosurveillance applications like the Defense Threat Reduction Agency (DTRA) Biosurveillance Ecosystem (BSVE) [11].

By focusing on these areas, we demonstrate TrustLab$^{\text{TM}}$ as a robust tool for identifying trusted online information sources across varied domains. The following sections outline TrustLab$^{\text{TM}}$'s capabilities and potential use cases.

C. Topic Selection and Trust Attributes Distribution

In the "Topics" section of TrustLab$^{\text{TM}}$ shown in Figure 1, we present comprehensive trust scores for each user, categorized by specific topics such as elections, the stock market, or various diseases. This section is broken down as follows: The "TrustLab$^{\text{TM}}$ Trust Filters" table (Table in Figure 1) ranks users by any targeted trust attribute or by the average score across all trust attributes, offering a nuanced view of user trustworthiness within specific contexts.

The "Trust Score Linear Distribution" chart (Figure 2a) illustrates the distribution of users across each trust attribute, with scores ranging from 0 to 1. This visual representation allows us to swiftly grasp the range and diversity of trustworthiness across different topics, thereby highlighting the spectrum of user credibility.

Furthermore, the "Trust Filter Score Per Source" chart (Figure 2b) details the composition of trust scores for the most or least trusted users. This facilitates an easy examination of the traits that distinguish highly trusted users from those deemed less reliable, thereby offering insights into the factors that contribute to a user's perceived trustworthiness.

Lastly, the "Source Network" chart (Figure 2c) maps out the social media connections among users based on their interactions, such as retweets, replies, or likes.

This chart effectively reveals clusters of social interconnections, thereby identifying the influencers within the network. Together, these visual tools provide a robust framework for analyzing and understanding the dynamics of trust across various topics on social media.

This section is designed to offer users the flexibility to explore a wide range of combinations pertaining to target domains, trust attributes, levels of trust, and intricacies of social network interconnections. For instance, researchers interested in delving deeper into the subject could use the distribution of trust attributes as a starting point to identify potentially significant attributes. Following this, TrustLab$^{\text{TM}}$ users can examine these specific trust attributes in more detail within the table, comparing them across different target domains to gain insight into their relative importance and application in various contexts.

In addition, we provide access to the original, unprocessed data for users who wish to conduct a thorough analysis of the results. This feature is especially valuable for those looking to explore raw data for more nuanced insights or to apply their own methodologies for data analysis.

In subsequent sections, we introduce other tools that offer refined results, thereby enabling users to efficiently leverage the TrustLab$^{\text{TM}}$ platform for a variety of purposes. These tools are designed to simplify the process of analyzing trust within social networks, making it more accessible to users to apply the platform's insights to their research or practical applications.

D. Sources, Users, and Trustworthiness Assessment

1) Comparison: The "Comparison" tab in the TrustLab$^{\text{TM}}$ interface enables users to evaluate the trustworthiness of various topics and information sources, helping them make informed decisions about online information.

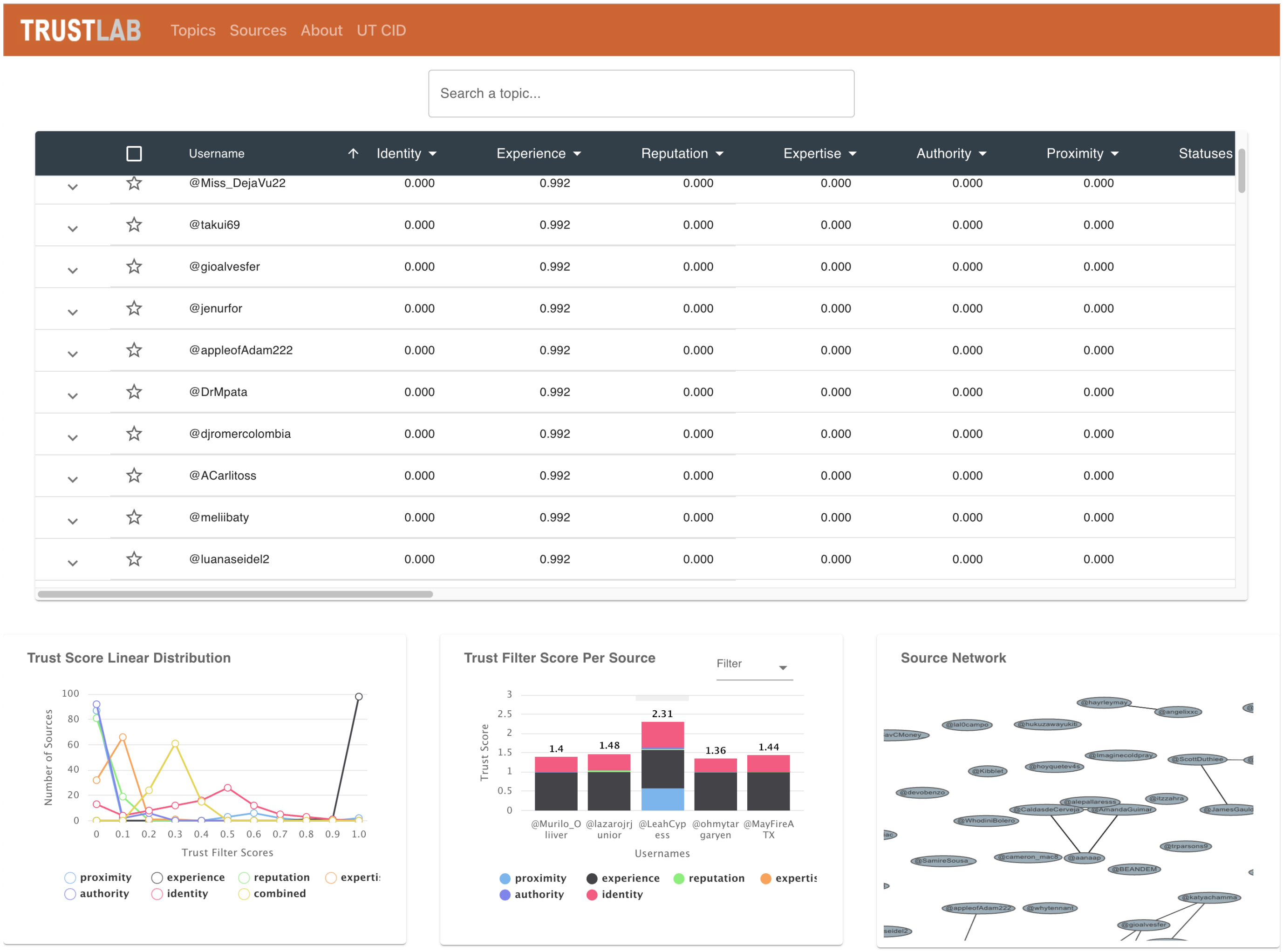
Users start by selecting a topic and a source (e.g., "Flu" and "CDC MMWR Quick Stats") as shown in Figure 3. Additional sources, like "WHO - Disease News", "BIOFEEDS HealthMap", and "WHER Reports", can be added to the comparison. Sources appear as green dots on the graph, and users can hover over these dots to view details or remove them with a click.

TrustLab$^{\text{TM}}$ also allows automatic comparison with the five most and least trusted sources (see Figure 4). By selecting specific Trust Filters, such as "Experience", users can see how sources compare based on individual or combined trust scores, making it easy to identify more trustworthy sources at a glance.

2) Recommendation: The "Score" section offers a detailed visualization of the trust scores for each trust attribute related to a specific user within a chosen topic. This allows for a direct comparison of an individual's trust scores with those of typical users and experts, including the variance observed within each group (Figure 5). This comparative analysis is based on the trust attribute extraction and group classification methodologies outlined by Huang et al. [9], [10].

Recommendations for enhancing a user's trustworthiness were derived from the findings of Huang et al. [9]. This research investigates strategies through which typical social media users might adjust their behaviors to bolster their trustworthiness and achieve recognition as experts in specific topics. To tailor these recommendations, a basin-hopping optimization algorithm was introduced by Wales et al. [15], is employed. This algorithm, known for its efficacy in global optimization, especially in complex, high-dimensional landscapes, applies

Copyright (c) IARIA, 2025.   ISBN: 978-1-68558-236-4



TRUSTLAB Topics Sources About UT CID

Search a topic...

	□	Username ↑	Identity ▾	Experience ▾	Reputation ▾	Expertise ▾	Authority ▾	Proximity ▾	Statuses
⌄	☆	@Miss_DejaVu22	0.000	0.992	0.000	0.000	0.000	0.000	
⌄	☆	@takui69	0.000	0.992	0.000	0.000	0.000	0.000	
⌄	☆	@gioalvesfer	0.000	0.992	0.000	0.000	0.000	0.000	
⌄	☆	@jenurfor	0.000	0.992	0.000	0.000	0.000	0.000	
⌄	☆	@appleofAdam222	0.000	0.992	0.000	0.000	0.000	0.000	
⌄	☆	@DrMpata	0.000	0.992	0.000	0.000	0.000	0.000	
⌄	☆	@djromercolombia	0.000	0.992	0.000	0.000	0.000	0.000	
⌄	☆	@ACarlitoss	0.000	0.992	0.000	0.000	0.000	0.000	
⌄	☆	@meliibaty	0.000	0.992	0.000	0.000	0.000	0.000	
⌄	☆	@luanaseidel2	0.000	0.992	0.000	0.000	0.000	0.000	

Fig. 1: Overview of Topics Page.

random perturbations to transcend local minima and employs a local search algorithm to refine solutions within each basin.

By comparing a specific user's trust scores with those of a typical user group and experts, the platform facilitates clear understanding of the disparities between them. Consequently, tailored trust score recommendations offer actionable guidance for users aiming to enhance their perceived trustworthiness on social media. This feature not only aids in personal or professional development but also contributes to the broader goal of fostering a more trustworthy and reliable digital community.

E. Use Cases

In this section, the TrustLab^{TM} use cases provided in Table I demonstrate potential application scenarios and utility for online users.

III. TrustLab^{TM} and a Comparison to Related Work

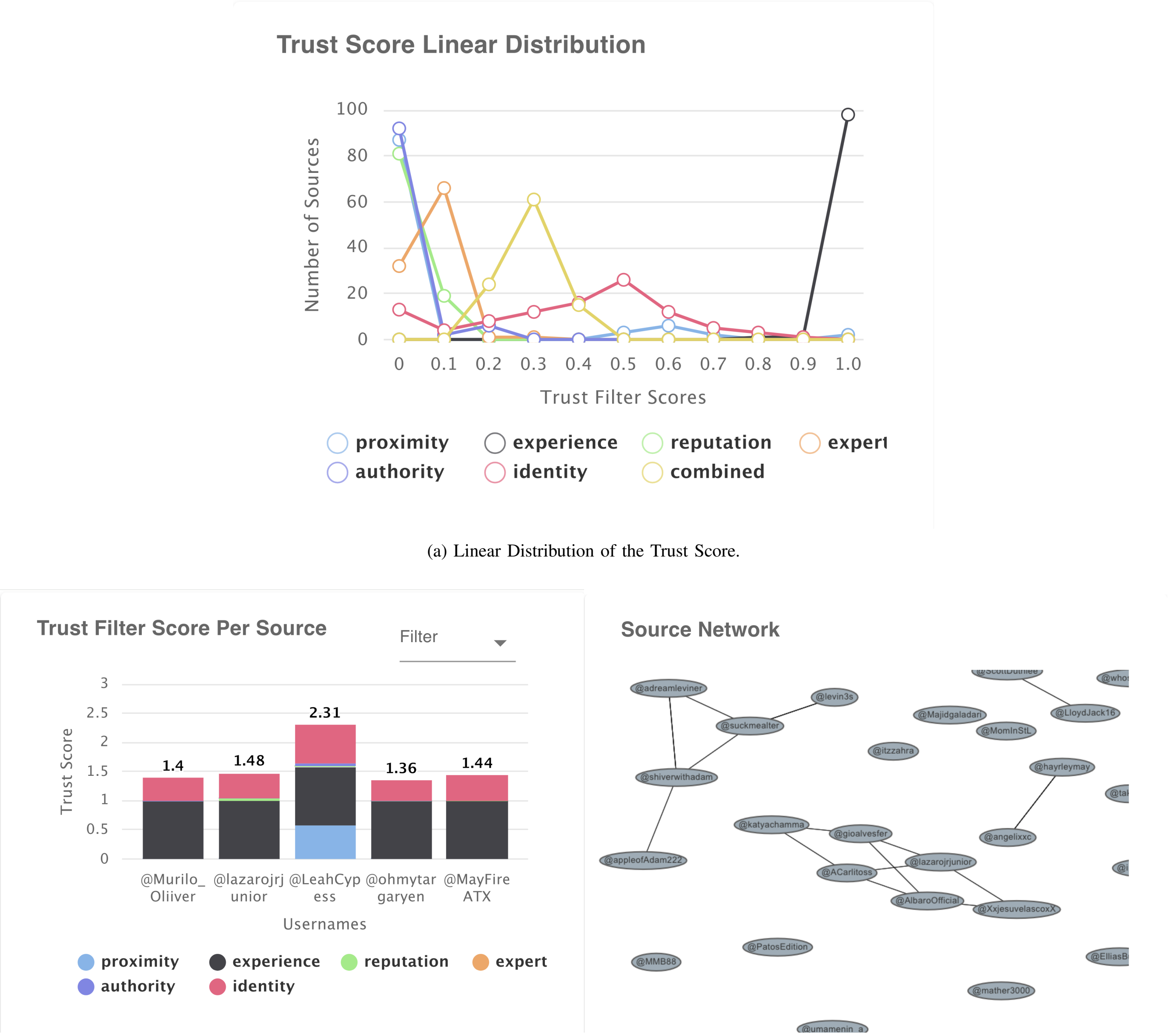
With the rise of social media as a primary source of news and information for many people, there is growing concern about the accuracy and reliability of the content shared on these platforms. Information trustworthiness on social media platforms is a critical issue that has attracted considerable research effort. Many systems and tools have also been developed to detect and counter misinformation, disinformation, or rumors.

Some studies have analyzed the content shared to assess its accuracy, bias, and potential for misinformation or disinformation. Using our tool, we assessed the trustworthiness of individuals sharing information. In this section, TrustLab^{TM} is compared with seven state-of-the-art trust-related social media tools.

In Table II, TrustLab^{TM} is compared with seven state-of-the-art trust-related tools highlighting the differences between these tools and discussing the pros and cons of each tool's technology, capabilities, and methods.

In Table III, several comparison metrics were selected to compare the representative tool features and functionality of these trust-related tools. First, we discuss the domains in which these tools have proven to be effective. Finance,

Copyright (c) IARIA, 2025. ISBN: 978-1-68558-236-4



(a) Linear Distribution of the Trust Score.

(b) Trust Score Per Source Graph.

(c) Source Network Graph.

Fig. 2: Topics Page (Figure 1) Breakdown.

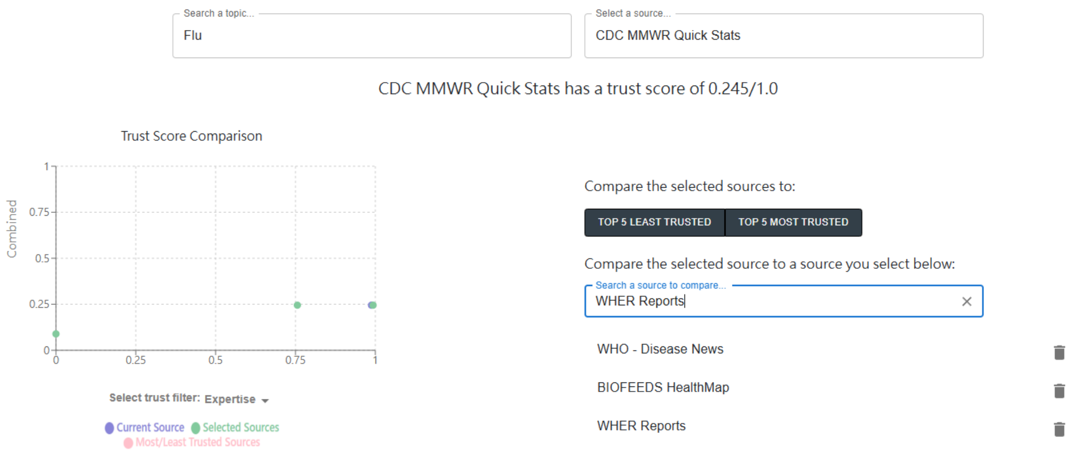
politics, and health are three of the most important domains, in which combating misinformation and disinformation is crucial. Some tools only perform experiments in a single domain or specific dataset, which limits their proven efficacy. Second, we compared the ability of the tool to differentiate between different objects. We evaluate the tool's ability to distinguish trusted information sources from malicious sources, trusted content from unreliable content, trusted users from untrusted users, or even bots. Third, we evaluate whether it is a platform-oriented tool or a user-oriented tool; in other words, whether it was developed to serve social media platforms such as X or whether it was developed to serve online users.

The comparison table highlights the distinct advantages of TrustLab^{TM} over other existing solutions in the field. TrustLab^{TM} excels for several reasons.

TrustLab^{TM} goes beyond focusing on a single target domain in social media to encompass multiple domains. This versatility allows it to be applied across diverse areas related to public opinion or sentiment, thereby providing a comprehensive solution for trust assessment in various contexts.

It also demonstrates exceptional capabilities in differentiating users, content, and sources based on their trustworthiness. This granular approach enables the nuanced evaluation of trust dynamics within online communities, thereby enhancing the reliability of trust assessments.

TrustLab^{TM} offers significant benefits to both users and



Search a topic... Flu

Select a source... CDC MMWR Quick Stats

CDC MMWR Quick Stats has a trust score of 0.245/1.0

Compare the selected sources to:

Compare the selected source to a source you select below:

Search a source to compare... WHER Reports

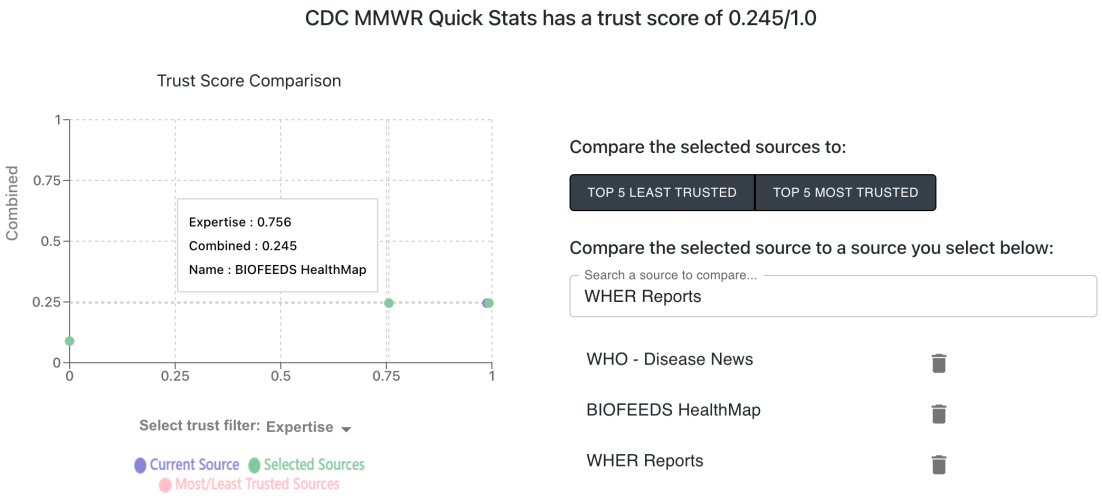
WHO - Disease News

BIOFEEDS HealthMap

WHER Reports

Select trust filter: Expertise

Fig. 3: Comparison Tab Sources Selection.



CDC MMWR Quick Stats has a trust score of 0.245/1.0

Expertise : 0.756
Combined : 0.245
Name : BIOFEEDS HealthMap

Compare the selected sources to:

Compare the selected source to a source you select below:

Search a source to compare... WHER Reports

WHO - Disease News

BIOFEEDS HealthMap

WHER Reports

Select trust filter: Expertise

Fig. 4: Comparison Tab Source Info.

platforms. Addressing the needs of both stakeholders fosters a more trustworthy and transparent online environment, promoting positive interactions and informed decision-making.

In conclusion, TrustLab^{TM}'s multifaceted approach, combined with its user-centric design and platform integration, positions it as a leading tool for trust assessment in social media across various domains.

IV. Conclusion and Future works

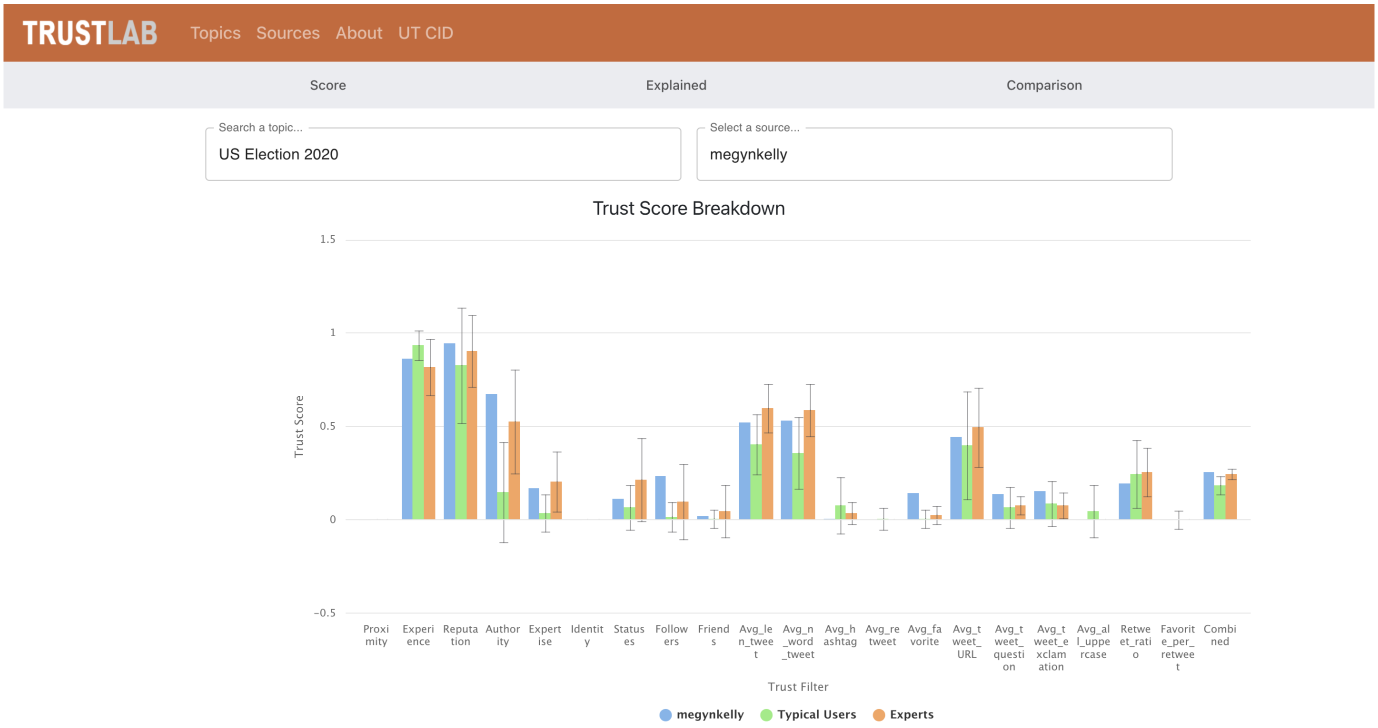
TrustLab^{TM} is an effective tool for evaluating online trustworthiness, significantly enhancing the accuracy and reliability of digital interactions across various domains such as finance, politics, and health. By providing a clear metric to distinguish between credible and less trustworthy sources, TrustLab^{TM} empowers users to make informed assessments of the information and information sources on which they rely to make decisions, fostering a safer and more transparent online environment. Moreover, the inclusion of personalized recommendations to improve individual trustworthiness is a standout feature of TrustLab^{TM}. These recommendations not only guide users in enhancing their own online presence but also contribute to the overall trustworthiness of digital communities by elevating the quality of interactions and information exchange. This dual capability of assessing and improving trust makes TrustLab^{TM} a vital tool in the pursuit of more reliable and ethical online engagements.

Looking forward, the development roadmap for TrustLab^{TM} includes several promising enhancements:

1) Customization of Trust Attributes: Enabling users to define or adjust trust attributes allows TrustLab^{TM} to meet specific contextual needs, making it versatile across different platforms and user requirements.
2) Enhanced Integration and Accessibility: By improving integration with other tools and refining the user interface, TrustLab^{TM} aims to become more accessible to a broader audience, including those with limited technical expertise, thus expanding its utility and effectiveness.

These initiatives aim to further cement TrustLab^{TM}'s role as a cornerstone technology for enhancing the integrity and relia-

Copyright (c) IARIA, 2025. ISBN: 978-1-68558-236-4



Search a topic...
US Election 2020

Select a source...
megynkelly

Fig. 5: Trust Score Breakdown.

TABLE I: USE CASES OF TrustLab$^{\text{TM}}$ TRUST FILTERS.

Use Scenarios	**Application Notes**
To study online trust indicators	The Trust Score Table (Figure 1) shows calculated trust scores of X users on seven trust attributes: Authority, Experience, Expertise, Identity, Proximity, and Reputation.
To analyze trust score distribution for a topic	The Trust Score Linear Distribution Chart (Figure 1) displays trust score distribution for a topic. Low trust among most sources may indicate misinformation requiring further research.
To examine relationships between information sources	The Source Network Graph (Figure 1) represents sources as nodes and shows their connections, illustrating proximity and relevance.
To evaluate trustworthiness across topics and sources	The Trust Score Comparison Scatter Plot (Figure 4) identifies the top five most and least trustworthy sources. The Linear Distribution Chart also highlights overall trust levels for each topic.
To identify high- and low-quality sources on social media	The Trust Score Per Source Graph (Figure 1) presents stacked trust scores for each attribute, identifying the most and least trustworthy sources.
To obtain a source's trust score on specific attributes	The Trust Score Explained page (Figure 6) details trust scores by attribute, helping users make informed decisions.
To understand how trust is built online	The About pages (Figure 7) explain the computation of the seven trust attributes, giving an overview of how trustworthiness is measured.
To receive advice on improving trust scores and influence	The Trust Score Breakdown page (Figure 5) offers suggestions for increasing social media credibility and reach within specific fields.

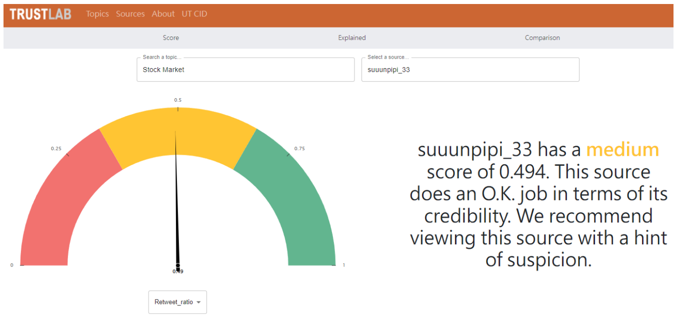
TABLE II: TECHNICAL COMPARISON WITH TRUST-RELATED TOOL COUNTERPARTS FOR SOCIAL MEDIA.

Trust Tools for social media	**Technology**	**Capabilities**	**Methods**
TrustLab$^{\text{TM}}$ Trust Filters	Random Forest Classification; Random Forest for Time Series Prediction; Basin Hopping Optimization	A tool to find trusted information on social media by filtering and scoring trusted users. Ability to classify user groups based on trustworthiness and provide recommendations on improving trustworthiness for users.	Quantify user trustworthiness under multiple trust attributes, and performs expert detection and trustworthy user ranking.
Bot Sentinel [16]	Machine Learning Classification	A platform that classifies and tracks inauthentic accounts and toxic trolls on X. Records marked accounts in a database.	Classify and scores accounts based on how likely the account engages in nefarious activities, which may result in the spread of disinformation.
Mendoza et al. [5]	PyTorchBigGraph (PBG); Proximity Graph; In-order Traversal	A semi-supervised algorithm to distinguish between bots and legitimate users. This work also examined the impact of malicious accounts on the spread of misinformation.	Demonstrate the existence of different robot clusters through label propagation and interaction graph analysis. Identified potential areas where misinformation could have spread.
Lukasik et al. [17]	Gaussian Process; Multi-task Learning; Natural Language Processing (NLP)	A transfer learning approach for classifying stances in tweets discussing emerging rumors.	Determine aggregate stance of a rumor, which has been shown to generally correlate with actual rumor veracity. Enables users to be more informed on the validity of rumors on X.
Gilani et al. [6]	Behavioral Analysis; Interaction Graphs	A comparative analysis of bots and legitimate users on Twitter (X). This work uncovers differences in account behavioral characteristics between bots and humans to facilitate bot detection.	Reveal the profound impact on the spread of information upon removing bots from X. Although bots are a major factor in the spread of misinformation and rumors, the detriment on the overall spread of any information may outweigh the benefits of removing bots.
SENTINEL [18]	Machine Learning Classification; Deep Neural Networks	A software system to classify health-related tweets and detect disease outbreaks. It also provides instant predictions of current disease levels.	Combat misinformation in posts by validating them with information from other trustworthy news and data sources. Ensures users receive correct and verified information.
Nizzoli et al. [19]	User Similarity Network	A network-based framework for detecting coordinated behavior and discovering coordinated communities on social media. This work also characterizes the coordination patterns that emerge in different community behaviors.	Analyze the similarity and degree of coordination in posts. May give insight into potential organized misinformation or rumors. Unable to confidently determine if coordination was intentional or coincidental.
TrollPacifier [20]	Sentiment Analysis; ActoDeS Framework	A holistic system for troll detection of users on Twitter (X) with high accuracy.	Address potential disinformation by identifying accounts with potentially malicious behaviors.

TABLE III: FUNCTIONAL COMPARISON WITH TRUST-RELATED TOOL COUNTERPARTS FOR SOCIAL MEDIA.

Trust Tools for Social Media	**Target Domains**				**Ability to Distinguish**			**Developed to Serve**	
	Finance	Politics	Health	Other	Sources	Content	Users	Platform	Users
TrustLab$^{\text{TM}}$ Trust Filters	✔	✔	✔		✔	✔	✔	✔	✔
Bot Sentinel [16]						✔	✔	✔	✔
Mendoza et al. [5]				Music			✔	✔	
Lukasik et al. [17]				N/A		✔	✔	✔	
Gilani et al. [6]				N/A			✔	✔	
SENTINEL [18]			✔		✔	✔		✔	✔
Nizzoli et al. [19]		✔					✔	✔	
TrollPacifier [20]				N/A		✔	✔		✔

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org

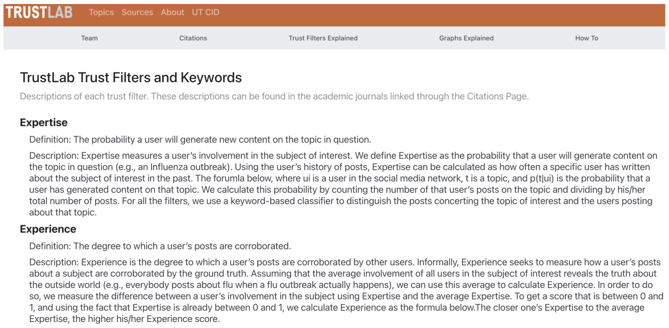


Search a topic... Stock Market

Select a source... suuunpipi_33

suuunpipi_33 has a medium score of 0.494. This source does an O.K. job in terms of its credibility. We recommend viewing this source with a hint of suspicion.

Fig. 6: Trust Score Explained.



TrustLab Trust Filters and Keywords

Descriptions of each trust filter. These descriptions can be found in the academic journals linked through the Citations Page.

Expertise

Definition: The probability a user will generate new content on the topic in question.

Description: Expertise measures a user's involvement in the subject of interest. We define Expertise as the probability that a user will generate content on the topic in question (e.g., an Influenza outbreak). Using the user's history of posts, Expertise can be calculated as how often a specific user has written about the subject of interest in the past. The forumla below, where ui is a user in the social media network, t is a topic, and p(t|ui) is the probability that a user has generated content on that topic. We calculate this probability by counting the number of that user's posts on the topic and dividing by his/her total number of posts. For all the filters, we use a keyword-based classifier to distinguish the posts concerning the topic of interest and the users posting about that topic.

Experience

Definition: The degree to which a user's posts are corroborated.

Description: Experience is the degree to which a user's posts are corroborated by other users. Informally, Experience seeks to measure how a user's posts about a subject are corroborated by the ground truth. Assuming that the average involvement of all users in the subject of interest reveals the truth about the outside world (e.g., everybody posts about flu when a flu outbreak actually happens), we can use this average to calculate Experience. In order to do so, we measure the difference between a user's involvement in the subject using Expertise and the average Expertise. To get a score that is between 0 and 1, and using the fact that Expertise is already between 0 and 1, we calculate Experience as the formula below.The closer one's Expertise to the average Expertise, the higher his/her Experience score.

Fig. 7: TrustLab$^{\text{TM}}$ Trust Filters Explanation.

bility of online content, ensuring that it continues to contribute effectively to the trustworthiness of digital communication.

References

[1] I. Goodfellow *et al.*, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.

[2] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[3] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, and R. Procter, "Detection and resolution of rumours in social media: A survey," *Acm Computing Surveys (Csur)*, vol. 51, no. 2, pp. 1–36, 2018.

[4] C. Naumzik and S. Feuerriegel, "Detecting false rumors from retweet dynamics on social media," in *Proceedings of the ACM web conference 2022*, 2022, pp. 2798–2809.

[5] M. Mendoza, M. Tesconi, and S. Cresci, "Bots in social and interaction networks: detection and impact estimation," *ACM Transactions on Information Systems (TOIS)*, vol. 39, no. 1, pp. 1–32, 2020.

[6] Z. Gilani, R. Farahbakhsh, G. Tyson, and J. Crowcroft, "A large-scale behavioural analysis of bots and humans on twitter," *ACM Transactions on the Web (TWEB)*, vol. 13, no. 1, pp. 1–23, 2019.

[7] G. Lingam, R. R. Rout, D. V. Somayajulu, and S. K. Das, "Social botnet community detection: a novel approach based on behavioral similarity in twitter network using deep learning," in *Proceedings of the 15th ACM Asia Conference on Computer and Communications Security*, 2020, pp. 708–718.

[8] H. Peng *et al.*, "Unsupervised social bot detection via structural information theory," *arXiv preprint arXiv:2404.13595*, 2024.

[9] T.-C. Huang and K. S. Barber, "Using historical social media retrieved trust attributes to help distinguishing trustworthy users," *INTELLI 2023*, p. 21, 2023.

[10] T.-C. Huang, R. N. Zaeem, and K. S. Barber, "Identifying real-world credible experts in the financial domain," *Digital Threats: Research and Practice*, vol. 2, no. 2, pp. 1–14, 2021.

[11] R. N. Zaeem, D. Liau, and K. S. Barber, "Predicting disease outbreaks using social media: Finding trustworthy users," in *Proceedings of the Future Technologies Conference (FTC) 2018: Volume 1*. Springer, 2019, pp. 369–384.

[12] T.-C. Huang, R. N. Zaeem, and K. S. Barber, "It is an equal failing to trust everybody and to trust nobody: Stock price prediction using trust filters and enhanced user sentiment on twitter," *ACM Transactions on Internet Technology (TOIT)*, vol. 19, no. 4, pp. 1–20, 2019.

[13] G. Lin, R. N. Zaeem, H. Sun, and K. S. Barber, "Trust filter for disease surveillance: Identity," in *2017 Intelligent Systems Conference (IntelliSys)*. IEEE, 2017, pp. 1059–1066.

[14] T.-C. Huang, R. N. Zaeem, and K. S. Barber, "Finding trustworthy users: Twitter sentiment towards us presidential candidates in 2016 and 2020," in *Intelligent Systems and Applications: Proceedings of the 2021 Intelligent Systems Conference (IntelliSys) Volume 2*. Springer, 2022, pp. 804–821.

[15] D. J. Wales and J. P. Doye, "Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms," *The Journal of Physical Chemistry A*, vol. 101, no. 28, pp. 5111–5116, 1997.

[16] C. Bouzy, "Bot sentinel," https://botsentinel.com/, [retrieved: Jan, 2025].

[17] M. Lukasik *et al.*, "Gaussian processes for rumour stance classification in social media," *ACM Transactions on Information Systems (TOIS)*, vol. 37, no. 2, pp. 1–24, 2019.

[18] O. Șerban, N. Thapen, B. Maginnis, C. Hankin, and V. Foot, "Real-time processing of social media with sentinel: A syndromic surveillance system incorporating deep learning for health classification," *Information Processing & Management*, vol. 56, no. 3, pp. 1166–1184, 2019.

[19] L. Nizzoli, S. Tardelli, M. Avvenuti, S. Cresci, and M. Tesconi, "Coordinated behavior on social media in 2019 uk general election," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 15, 2021, pp. 443–454.

[20] P. Fornacciari, M. Mordonini, A. Poggi, L. Sani, and M. Tomaiuolo, "A holistic system for troll detection on twitter," *Computers in Human Behavior*, vol. 89, pp. 258–268, 2018.