Truong et al. *EPJ Data Science* (2024) 13:10
https://doi.org/10.1140/epjds/s13688-024-00450-9

EPJ Data Science
a SpringerOpen Journal

EPJ.org

RESEARCH Open Access

Account credibility inference based on news-sharing networks

Bao Tran Truong[1]*, Oliver Melbourne Allen[1,2] and Filippo Menczer[1]

*Correspondence: baotruon@iu.edu
[1]Observatory on Social Media, Indiana University, Bloomington, USA
Full list of author information is available at the end of the article

Abstract

The spread of misinformation poses a threat to the social media ecosystem. Effective countermeasures to mitigate this threat require that social media platforms be able to accurately detect low-credibility accounts even before the content they share can be classified as misinformation. Here we present methods to infer account credibility from information diffusion patterns, in particular leveraging two networks: the reshare network, capturing an account's trust in other accounts, and the bipartite account-source network, capturing an account's trust in media sources. We extend network centrality measures and graph embedding techniques, systematically comparing these algorithms on data from diverse contexts and social media platforms. We demonstrate that both kinds of trust networks provide useful signals for estimating account credibility. Some of the proposed methods yield high accuracy, providing promising solutions to promote the dissemination of reliable information in online communities. Two kinds of homophily emerge from our results: accounts tend to have similar credibility if they reshare each other's content or share content from similar sources. Our methodology invites further investigation into the relationship between accounts and news sources to better characterize misinformation spreaders.

Keywords: Information spread; Social media; Misinformation; Network analysis; Credibility

1 Introduction

Many people are now getting news from social media [1]. With just a click on the "Share" button, anyone can be a broadcaster of news on these platforms. Such a low barrier, combined with uneven journalistic standards in online news, has facilitated the spread of misinformation in the information ecosystem [2]. Such proliferation of misinformation poses grave threats to democracy [3], the economy [4], and public health [5–8].

Existing methods to curb misinformation focus on classifying either the content or the account posting it. However, multiple challenges exist for content-based methods, posing a need for methods to evaluate sources instead of (or in addition to) the content itself. In particular, traditional fact-checking methods involving human efforts to manually verify the accuracy of claims cannot scale with the sheer volume and speed of information being shared online. Automatic misinformation detection could potentially overcome the problem of scale. But when they work, these methods rely on extensive language features such

Springer

as writing style, lexicon, and emotion [9]. Therefore they need to be constantly retrained to reflect new knowledge and evolving tactics employed by purveyors of false information. These challenges are exacerbated by the rise of content created by generative AI. The availability of open-source large language models (LLMs) brings down the cost of generating content, creating opportunities for malicious actors to spread misleading content and influence public opinion [10, 11]. Worse yet, LLMs' ability to mimic writing styles makes this type of content persuasive yet very challenging to detect [12].

In this paper, we propose several methods to infer the credibility of news-sharing accounts on social media to detect low-credibility accounts likely to spread misinformation. The credibility of a news source or account may depend on many factors, including reputation and adherence to factual reporting and transparency. More precisely, we define *high-credibility news sources* to be those that meet objective journalistic criteria as assessed by third-party fact-checking organizations [13, 14]. Following that, *credible accounts* are those who share or reshare high-credibility sources. Credibility indicators provide useful signals for consumers to navigate the information landscape: they not only help people seeking information outlets [15] but can also decrease the propensity to share misinformation [16] by raising user awareness. Knowing unreliable, influential accounts might also aid platforms in mitigating their impact.

News-sharing decisions on social media depend on the actual content as well as on *trust* in *who* shares it [15, 17] and *what* media outlet it originates from [18]. When information sharing on social media is represented as a network connecting accounts and/or news sources, we can apply network analysis methods to infer node properties by propagating labels across links [19–21]. Most existing network-based methods to detect low-credibility accounts [20] focus solely on interactions between accounts. However, bipartite networks capturing the reinforcing relationship between news outlets and consumers have been shown to be effective for classifying misinformation content [22]. This merits further examination of the interactions between accounts and sources to better understand the characteristics of misinformation spreaders.

Social context is useful in detecting fake news [23]. Generalizing this observation, we hypothesize that it is possible to infer the credibility of an account by looking at either the sources or the other accounts they trust. The fundamental assumption underlying such inference is the existence of *credibility homophily* among misinformation spreaders. Homophily can be defined based on different network relationships, such as following, resharing, co-sharing, and so on. For example, adjacent nodes in the reshare network (accounts that reshare each other) trust each other, so they should have similar credibility. While *trust in accounts* and *trust in sources* are both reasonable assumptions for the task, no other work has compared their effectiveness in the same settings.

We explore three research questions:

- **Q1:** Is there credibility homophily in the reshare network?
- **Q2:** Is there credibility homophily in the co-share network?
- **Q3:** Can we infer the credibility of accounts by leveraging such homophily and looking at neighbor nodes in the reshare or bipartite/co-share networks, i.e., by examining the trust relationships among accounts or between accounts and sources?

In this paper, we propose several network-based methods — including centrality measures and graph embeddings — to infer the credibility of news-sharing accounts on social media. We explore **Q1** using the reshare network, where a directed edge represents trust

by one account in another. We explore **Q2** using the bipartite network where accounts are connected to the news sources they share. The paper makes three contributions:

1. We introduce several methods to measure the credibility of accounts by leveraging different kinds of information-sharing networks.
2. We introduce an evaluation framework to systematically estimate and compare the accuracy of our algorithms using empirical data from multiple contexts and social media platforms.
3. We show that there are two kinds of homophily among information spreaders: accounts tend to reshare content from individuals with similar credibility (**Q1**) and to share content from the same sources as individuals with similar credibility (**Q2**). These diffusion patterns explain the effectiveness of network methods that estimate an account's credibility using their trust in other accounts or in sources (**Q3**).

After reviewing related work, we present our methodology, including the definition of the credibility inference task and algorithms leveraging an account's trust in other accounts and in sources, respectively. We then describe the experimental setup and discuss the evaluation results.

2 Related work

One approach to detect credible accounts uses heuristics such as the assumption that online opinion leaders are credible [24]. This assumption is violated by misinformation superspreaders [8, 25, 26] and leaves out potentially credible ordinary users. Therefore this paper focuses on network-based approaches that leverage accounts' connectivity in addition to heuristics.

PageRank [27] is a widely used centrality measure that assigns scores for nodes in a directed network by simulating a diffusion process through the network analogous to random surfing among web pages. *Personalized PageRank* [28] incorporates prior knowledge about the importance of some nodes by constraining the random surfer to stay close to those nodes. Methods extending PageRank and Personalized PageRank have been applied to measure trust in peer-to-peer (P2P) networks. *EigenTrust* was used to obtain global reputation values for each user [29]; *PowerTrust* [30] and *TrustRank* [31, 32] were used to rank search results. Different from existing work, one of our proposed methods attempts to infer account credibility by finding the highest-ranking nodes in a network where misinformation, rather than trust, propagates. Methods have been introduced that similarly model the spread of "distrust" to measure trustworthiness in P2P networks [33–35]. To our knowledge, no prior work has explored distrust in social news-sharing networks.

Another well-known network centrality method is Hyperlink-Induced Topic Search (*HITS*) [36]. This method ranks the web pages returned by a search query by assuming that hubs are useful in leading a web surfer to authoritative pages. Several algorithms extend HITS. *Co-HITS* [37] and *Bipartite Graph Reinforcement Model (BGRM)* [38] incorporate pre-existing information about the relevance of some web pages to constrain the final scores. *BiRank* [39] is a similar extension developed for n-partite graphs. In the context of social media, HITS has been used to find influential users [40] as well as high-quality content [41]. One of our proposed methods extends HITS while maximizing prior knowledge about accounts to produce accurate credibility scores. The intuition is that misinformation sharers are hubs leading to unreliable news sources, and vice versa.

Machine learning methods have been employed to classify social media accounts based on their credibility. Previous studies have trained models on features engineered from

a user profile [42, 43] or the content they post [44, 45]. Graph embedding methods are popular means to obtain a compact representation of nodes in a network. Since network structure information is preserved, nodes with similar positions in the network have vectors that are close together in the embedding space [46–48]. These methods have been applied to classify rumor-spreading accounts in retweet networks, in combination with additional features such as an account's inferred believability [20], screen name, profile description, and activity level [49]. We explore simpler methods that only use vectors capturing an account's position in the network in some of the algorithms proposed here.

3 Methods

The task is formalized as follows: given (i) a set of posts with links to news articles and (ii) credibility labels for a subset of accounts, assign credibility scores to the unlabeled accounts.

An overview of the pipeline for mining social media for this task is presented in Fig. 1. The pipeline consists of five steps:

1. Social media data is collected and cleaned. The data includes a set of known labels and a list of interactions between accounts, for example, via tweets or retweets.
2. Different networks can be constructed to capture distinct potential signals. The reshare network captures an account's trust in other accounts. Some of our algorithms use the trust network — the transpose of this reshare network. The bipartite account-source network captures an account's trust in sources. The co-share network is obtained by projecting the bipartite network onto account nodes.
3. We label a subset of account nodes in the network. These labels are derived from source credibility scores provided by a fact-checking organization, such as NewsGuard[1] or Media Bias Fact Check.[2] To obtain an account label, we calculate

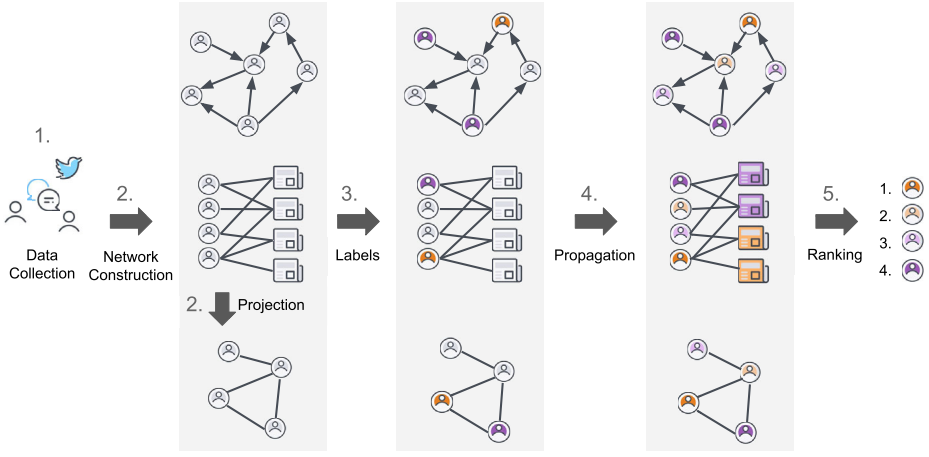

Figure 1 *Pipeline for mining trust patterns from information-sharing data.* Trust in accounts is modeled by the reshare network (top); trust in sources can be modeled by the bipartite account-source network (middle), or a projection of it, the co-share networks (bottom). Colors represent credibility labels: orange and purple for high- and low-credibility, respectively

[1] newsguardtech.com.

[2] mediabiasfactcheck.com.

the weighted mean of all the sources they shared and then apply a threshold to this mean score, following the organization's standard for the credibility threshold.

4. Each algorithm is applied to propagate or compute credibility scores for all nodes.
5. Lastly, unlabeled accounts are ranked by their credibility scores for evaluation.

In the next subsections, we present several methods to infer account credibility based on trust in accounts or sources. In each case, we evaluate centrality-based methods and an embedding algorithm. Centrality-based methods are less sophisticated but more efficient and interpretable.

3.1 Trust in accounts

We describe algorithms to calculate an account's credibility by leveraging their trust in other accounts. Trustworthiness can be inferred from the *trust network* G^T, in which a link goes from *Alice* → *Bob* if Alice follows or reshares Bob. In line with previous work, these actions can be considered endorsements that signal trust by the sharing account in the account being shared [20, 50–52].

The trust network is the transpose of the *reshare network* G, a weighted, directed graph where the nodes represent accounts and edges correspond to reshare interactions among accounts. Edges follow the direction of information spread (from reshared to resharing account) and are weighted by the number of reshares. Let $G_{ij} = n$ if j retweets i n times.

Finally, we assume that some accounts have credibility labels. Let H be a set of nodes that are known to have high credibility and L a set of nodes known to have low credibility.

3.1.1 PageRank Trust

The PageRank family of algorithms can be used to calculate account trustworthiness scores based on this signal. An account's *PageRank Trust* score is calculated iteratively using a weighted version of PageRank:

$$\tau_i^{t+1} = (1-\alpha)\sum_j \frac{G^T_{ji}}{\sum_\ell G^T_{j\ell}}\tau_j^t + \frac{\alpha}{N}, \tag{1}$$

where t is the iteration step, α is the teleportation factor. The intuition of this method is that accounts with many incoming links are trusted and therefore have high PageRank Trust scores, indicating high credibility.

3.1.2 Personalized PageRank Trust

Information about some high-credibility nodes may be available. To incorporate this information, an account's Personalized PageRank Trust is calculated as follows:

$$\tau_i^{t+1} = (1-\alpha)\sum_j \frac{G^T_{ji}}{\sum_\ell G^T_{j\ell}}\tau_j^t + \alpha\tau_i^0, \tag{2}$$

where τ_i^0 is defined by:

$$\tau_i^0 = \begin{cases} \frac{1}{|H|} & \text{if } i \in H, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Under this scheme, accounts that are trusted by credible accounts (H list) have higher Personalized PageRank Trust, indicating high credibility.

3.1.3 TrustRank

This method first creates a quality seed set, ideally to be evaluated by experts, then uses this list to propagate scores. To apply the method in our context, PageRank Trust (Eq. (1)) is first used to select a set S of seed accounts with highest PageRank Trust scores, where $|S|$ is a parameter. These seeds are labeled as "good" or "bad" depending on available credibility labels:

$$\tau_i^0 = \begin{cases} 1 & \text{if } i \in H \cap S, \\ 0 & \text{if } i \in L \cap S, \\ \frac{1}{2} & \text{otherwise.} \end{cases} \tag{4}$$

TrustRank scores are then calculated using Personalized PageRank Trust (Eq. (2)) with these τ_i^0 values.

3.1.4 LoCred

The methods mentioned so far rely on the observation that good pages seldom point to bad ones [31]. Decisions to circulate news online are not as straightforward — bad content is designed to mislead and people could be socially motivated. Therefore, we cannot assume that "good people seldom trust bad ones," or be sure that credible accounts do not reshare from low-credibility accounts.

We propose a method that uses the diffusion of information to capture an account's credibility without the assumption that accounts have good judgment. This method is performed on the reshare network. If a node reshares a lot, it might be gullible. These accounts might not spread misinformation intentionally but should be distrusted nonetheless. The Low Credibility Account Estimation (*LoCred*) score is devised to capture such spreaders of misinformation.

The LoCred score s_i for each node i is calculated iteratively until convergence as a weighted version of personalized PageRank:

$$s_i^{t+1} = (1-\alpha) \sum_j \frac{G_{ji}}{\sum_\ell G_{j\ell}} s_j^t + \alpha s_i^0, \tag{5}$$

where t is the iteration step, α is the teleportation factor, and the initial value s_i^0 is defined as follows:

$$s_i^0 = \begin{cases} \frac{1}{|L|} & \text{if } i \in L, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Accounts close to nodes in L have higher LoCred scores, indicating low credibility. Note that the initial values in Eqs. (3) and (6) are different. But even if they were the same, the rankings based on Eqs. (2) and (5) would not simply be the reverse of each other because the eigenvectors of a matrix and those of its transpose are different for asymmetric matrices.

3.1.5 Reputation scaling

Lastly, we introduce *Reputation Scaling*, a measure that balances the perceived trustworthiness of an account and its spreading of misinformation. This method builds on an application for distributed P2P systems [33].

The reputation of an account is "scaled" by combining both its LoCred and Personalized PageRank Trust score. We calculate both of these scores, then rank accounts based on a reputation score calculated as follows:

$$r_i = \tau_i(1 - s_i), \tag{7}$$

where τ_i is the trust of i calculated with Eq. (2) and s_i is its LoCred score calculated with Eq. (5). An account with high reputation is considered credible.

3.1.6 Node2vec on reshare network

We investigate whether an account's network position provides predictive signals about its credibility, i.e., accounts with similar network positions are similarly credible. To this end, we use node2vec [46] to obtain embedding vectors of accounts from the retweet network G. We use the typical values for node2vec parameters. Specifically, vector dimension, window size, number of walks, and walk length are 128, 10, 10, and 80, respectively. The optimization is run for ten epochs. Account node2vec vectors are then used in a k-Nearest Neighbors (KNN) classifier [53] with $k = 10$ to rank unlabeled accounts. We find the best results by performing a grid search over the random walk bias parameters $p, q \in \{0.25, 0.50, 1, 2, 4\}$. We evaluate each parameter combination using 5-fold cross-validation with 20% labeled data in each fold.

3.2 Trust in sources

We describe algorithms to calculate the credibility of accounts by leveraging their trust in sources (websites or domains in our case). We hypothesize that the mutually reinforcing relationship between news websites and news consumers can be applied in our bipartite account-source network to infer account credibility. Consider accounts as hubs and news domains as authorities: a "good" account is one pointing to many "good" —authoritative, high-credibility— websites; and vice versa, a good website is a node shared by many good accounts [36].

Let us define the *account-source network* as a weighted bipartite graph where nodes consist of two disjoint sets U and D that represent accounts and sources, respectively. A weighted edge represents the number of times an account shares links to a source. Formally, let G be the adjacency matrix such that $G_{ij} = n$ if account $i \in U$ shares links from source $j \in D$ n times. Similar to previous algorithms, we assume that some accounts have credibility labels.

3.2.1 HITS

HITS [36] updates account and source scores according to

$$u_i^{t+1} = \sum_{j \in D} A_{ij} d_j^t, \tag{8}$$

$$d_j^{t+1} = \sum_{i \in U} A_{ij} u_i^t, \tag{9}$$

respectively, where $A_{ij} = 1$ if $G_{ij} > 0$ and zero otherwise. The scores are normalized after each step t.

3.2.2 Co-HITS

Information about some high-credibility sources may be available. Co-HITS incorporates this information into the propagation to calculate node scores [37]. Account and source scores are calculated iteratively according to the strength of interactions between nodes from the two network partitions, until convergence:

$$u_i^{t+1} = \alpha u_i^0 + (1-\alpha) \sum_j \frac{G_{ij} d_j^t}{\sum_\ell G_{i\ell}}, \tag{10}$$

$$d_j^{t+1} = \beta d^0 + (1-\beta) \sum_i \frac{G_{ij} u_i^t}{\sum_\ell G_{\ell j}}, \tag{11}$$

respectively, where α and β are the teleportation factors for accounts and sources respectively, $d^0 = 1/|D|$ is the initial value for sources, and u_i^0 is the initial value for account i, defined as:

$$u_i^0 = \begin{cases} 0 & \text{if } i \in H, \\ 1 & \text{if } i \in L, \\ 1/|U| & \text{otherwise.} \end{cases} \tag{12}$$

These initial values are further normalized such that $\sum_{i \in U} u_i^0 = 1$. The scores are normalized, so no further normalization is necessary.

3.2.3 BGRM

BGRM [38] is similar to Co-HITS. The main difference is in the way they normalize node scores at each iteration (see Table 1 in [39] for more details).

3.2.4 BiRank

BiRank [39] is also similar to Co-HITS and BGRM, differing in the way node scores are normalized at each iteration (see Table 1 in [39] for more details).

3.2.5 CoCred

We propose Co-sharing network-based Credibility (*CoCred*), a measure that, like Co-HITS, BGRM, and BiRank, incorporates existing knowledge about some accounts' credibility into the credibility estimation of other accounts. The difference is that those algorithms apply the update rules to labeled accounts, whereas CoCred preserves the known labels, only rescaling them in the normalization step — we believe the accounts labels contain very strong signals.

The CoCred score u_i for each account i and the corresponding score d_j for each source j are updated at each timestep t according to

$$u_i^{t+1} = \begin{cases} u_i^0 & \text{if } i \in H \cup L, \\ \alpha u_i^0 + (1-\alpha) \sum_j \frac{G_{ij} d_j^t}{\sum_\ell G_{i\ell}} & \text{otherwise,} \end{cases} \tag{13}$$

$$d_j^{t+1} = \beta d^0 + (1-\beta) \sum_i \frac{G_{ij} u_i^t}{\sum_\ell G_{\ell j}}, \tag{14}$$

where α and β are the teleportation factors for accounts and sources, respectively; $d^0 = 1/|D|$ is the initial value for sources, and u_i^0 is the initial value for account i, as defined in Eq. (12). Note that Eq. (14) is the same as Eq. (11).

After each update, the scores are normalized so that $\sum_{i \in U} u_i = \sum_{j \in D} d_j = 1$. Our algorithm considers low-credibility accounts as those sharing unreliable news sources and vice versa. Low-credibility accounts have higher CoCred scores.

3.2.6 Node2vec on co-share network

The bipartite account-source network may also provide information about similar news-sharing patterns between accounts. To explore this, we project the bipartite network onto account nodes. Specifically, the *co-share network* is obtained by connecting accounts if they share links to the same sources. It is thus a weighted, undirected graph where the nodes are accounts having at least one shared source in common with another account. Each account is represented as a vector of shared domains. To mitigate the effect of popular sources on the similarity among accounts, we use Term Frequency — Inverse Document Frequency (TF-IDF) vector representations [54] where each dimension is a news source. Edges correspond to co-share interactions and are weighted by the cosine similarity between the TF-IDF vectors of the two connected accounts. Account embedding vectors obtained from the co-share network may reveal whether individuals who share information from the same sources have similar credibility. We use node2vec to obtain account vectors and use them in a KNN classifier with $k = 10$ to rank unlabeled accounts. The evaluation of this method uses the same parameter set and optimization procedure as that of node2vec on the reshare network described above.

4 Evaluation

Our evaluation framework only requires a set of social media posts, and not follower/friend data. We define the task of classifying low-credibility accounts as follows: given (i) a set of social media posts with links to news articles, and (ii) credibility labels for a subset of accounts, assign a binary label — *credible* or *not credible* — to the unlabeled accounts. The algorithms are evaluated on networks extracted from social media data. The datasets, corresponding networks, experimental setup, and results are described below.

4.1 Data

We consider three social media datasets.

- The Twitter_Covid dataset includes tweets about COVID-19, collected with the hashtags `#coronavirus` and `#covid19` from 9–29 March 2020 [55].
- The Twitter_Midterm dataset includes tweets containing hashtags and keywords about the U.S. 2022 Midterm elections [56]. We use a subset of this collection, from 8 October–18 November 2022.
- The Facebook_Midterm dataset contains Facebook posts collected the same way and spanning the same period as Twitter_Midterm [56].

We extract the data fields of interest from each social media post, including the user ID, post ID, and domain names of the shared links. If a link is shortened when shared

Table 1 Retweet network descriptions

Dataset	Nodes	Edges	Avg. degree	Credibility assortativity coeff.
Twitter_Covid	322,208	382,499	1.2	0.83
Twitter_Midterm	410,914	762,534	1.9	0.73

on these platforms, a pre-processing step is performed to extract the domain name from the expanded URL. Posts with domain names of platforms such as amazon.com, yelp.com, youtube.com, etc. are not considered. To limit noise, we further retain only accounts sharing at least five links and posts containing domains that are shared at least five times across the dataset. To assign scores to accounts, we start from source credibility scores, which are domain ratings obtained from NewsGuard in April 2021 for the Twitter_Covid and October 2022 for the Twitter_Midterm and Facebook_Midterm. We then compute an account score as a weighted mean of the scores of the sources they share. Note that not all accounts have a score as a result. Finally, following NewsGuard's rubric, we use a threshold of 60 for labeling accounts with a score below/above 60 as low/high-credibility.

4.2 Network description

4.2.1 Reshare network

Since reshare information is not included in Facebook data, the reshare network can only be constructed from the two Twitter[3] datasets out of the three datasets considered. The statistics of these retweet networks are summarized in Table 1. The table reports on network assortativity coefficients [57] using account credibility scores. The high coefficients are indicative of credibility homophily (answering **Q1**).

The homophily can also be observed in the core of the retweet network (Fig. 2), where low-credibility accounts tend to reshare from low-credibility accounts, and vice versa.

4.2.2 Bipartite network

The statistics of the bipartite networks are summarized in Table 2. The core of a bipartite network is visualized in Fig. 3.

4.2.3 Co-share network

Descriptive statistics of the co-share networks are presented in Table 3. There is an assortative mixing of credibility among accounts, suggesting the existence of credibility homophily in the co-share networks (answering **Q2**). The core backbone of the co-share network, shown in Fig. 4, illustrates this other type of homophily: similarly credible accounts tend to share content from similar sources.

4.3 Experimental setup

We employ the area under the receiver operating characteristic curve (ROC_AUC) and F1 score as metrics for low-credibility account classification. The maximum value of ROC_AUC, one, indicates that all low-credibility accounts are ranked before other accounts, whereas a value of 0.5 corresponds to random ranking. F1 is the harmonic mean of precision and recall. An F1 value of one means that the model correctly classifies all

[3] Because our data was collected from Twitter before it was renamed 'X,' we continue to refer to the platform by its original name.

Figure 2 *Retweet network constructed from the Twitter_Covid dataset.* To visualize the network, we filter nodes and edges using k-core decomposition [58] ($k = 3$). Colors represent credibility score: orange and purple for high- and low-credibility accounts, respectively, and gray for unlabeled accounts. Node size represents in-strength, i.e., the number of retweets by the account. Large purple nodes are known misinformation super-spreaders

Table 2 Bipartite network descriptions

Dataset	Accounts	Domains	Edges	Avg. degree
Twitter_Covid	474,094	55,539	613,609	1.3
Twitter_Midterm	126,445	13,415	1,190,390	9.4
Facebook_Midterm	6950	4141	26,252	17.1

classes; zero means all samples are wrongly classified. The F1 score requires converting the account credibility scores provided by the algorithms into binary labels using a threshold. We calculate precision and recall for a thousand threshold values spanning the unit interval and select the optimal threshold that maximizes the F1 score.

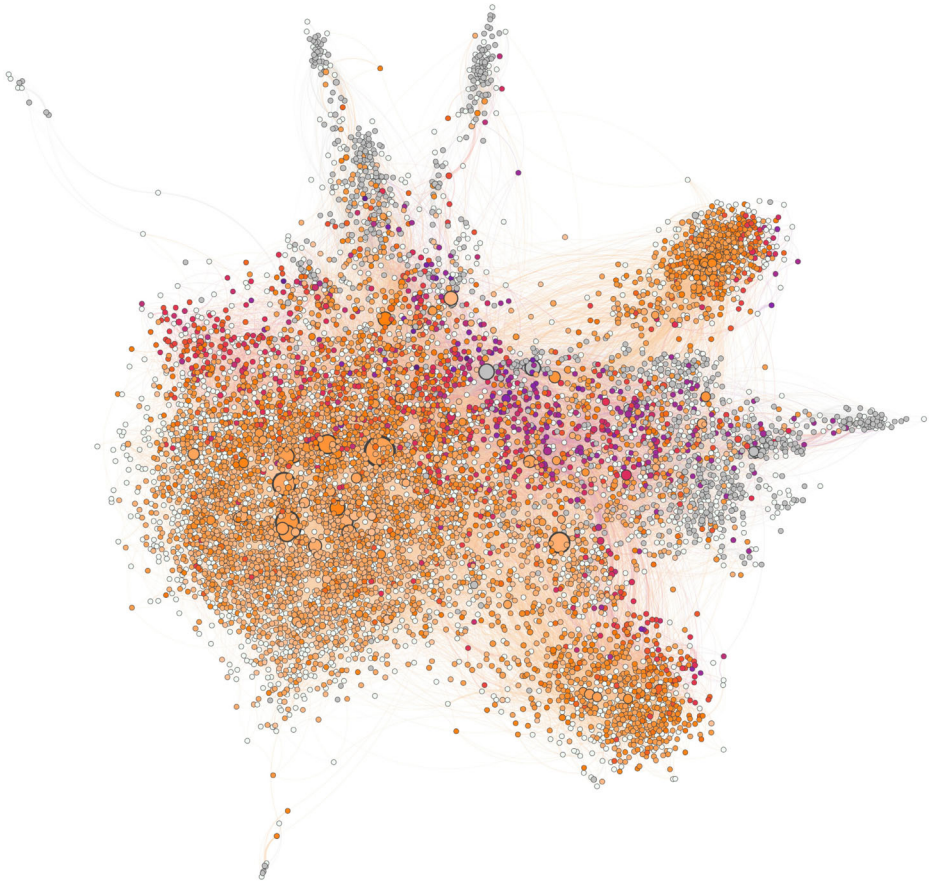

Figure 3 *Bipartite account-source network constructed from the Twitter_Covid.* For visualization, only the $k = 4$ core of the network is shown. Node colors represent account credibility scores: orange and purple for high- and low-credibility accounts, respectively, and gray for unlabeled accounts. Source nodes are in white. Node size represents the strength of an account node, i.e., the number of posts with links by that account

Table 3 Co-share network descriptions

Dataset	Nodes	Edges	Avg. degree	Credibility assortativity coeff.
Twitter_Covid	67,657	434,963	1441.7	0.5
Twitter_Midterm	115,461	59,491,099	515.3	0.8
Facebook_Midterm	3488	59,653	34.2	0.7

Since NewsGuard ratings are available for only a subset of domains shared in a dataset, we assign a *label confidence score* to an account by calculating the fraction of known domains that they shared; this number is between zero and one. The most stringent threshold is used in our evaluation — only accounts with a confidence score of one are used as *known accounts* to calculate evaluation metrics.

The evaluation uses 5-fold cross-validation. We hold out a random subset of 20% of known accounts (the test set). For centrality-based methods on the reshare and bipartite network, the initial scores of the remaining 80% of accounts are set to their true scores, while the test set starts with default scores at the beginning of propagation. Similarly, for classification using embeddings, we train the classifier on the remaining 80% of accounts.

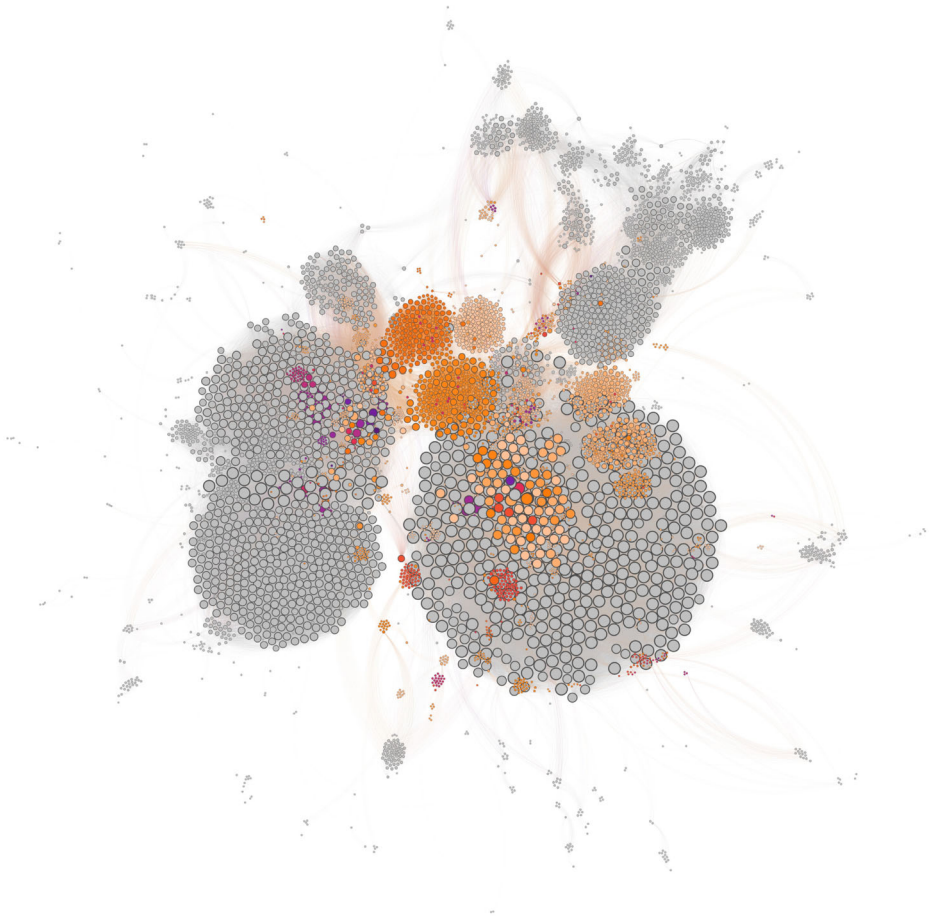

Figure 4 *Co-share network constructed from the Twitter_Covid*. Given the density of this network, we aggressively filter it for visualization. We first keep the 10% of edges with the highest weights and the 10% of nodes in the innermost core ($k = 4$). Then we apply the multiscale backbone method [59] (significance level 0.3) and retain the giant component of the backbone network. Colors represent credibility score: orange and purple for high- and low-credibility accounts, respectively; unlabeled accounts are in gray. Node sizes represent core numbers: smaller nodes are in the periphery of the network

Finally, for both cases, predicted labels for accounts in the test set are checked against the true labels. The reported metrics are averages of values across five folds.

We use $\alpha = \beta = 0.85$ for the teleportation factors in all network centrality algorithms. The best node2vec parameters for the retweet networks are $p = q = 1.0$. For the co-share network, the best node2vec parameters are $p = 2.0$, $q = 0.5$ for the Twitter_Covid dataset, $p = q = 1.0$ for the Twitter_Midterm dataset, and $p = 0.25$, $q = 4.0$ for the Facebook_Midterm dataset.

4.4 Results

Evaluation results for the low-credibility account classification task are summarized in Tables 4 (based on ROC_AUC) and 5 (based on F1 scores). The trends in performance are similar for both metrics, therefore we focus on ROC_AUC.

The results show that the trust relationships among accounts provide useful information to effectively infer their credibility. In particular, the best-performing method is node2vec on reshare networks, with a ROC_AUC score of more than 0.88 across datasets. The next

Table 4 Account classification ROC_AUC. Best results for each dataset are highlighted in darker shades

Sharing network	Method	Twitter_Covid	Twitter_Midterm	Facebook_Midterm
Reshare	node2vec	0.910 ± 0.006	0.885 ± 0.003	N/A
	LoCred	0.768 ± 0.004	0.773 ± 0.044	
	Rep. Scaling	0.618 ± 0.008	0.660 ± 0.003	
	Trustrank	0.534 ± 0.002	0.517 ± 0.001	
	PPR Trust	0.534 ± 0.002	0.517 ± 0.001	
	PR Trust	0.516 ± 0.002	0.520 ± 0.002	
Bipartite/Co-share	CoCred	0.802 ± 0.014	0.831 ± 0.013	0.715 ± 0.042
	node2vec	0.654 ± 0.020	0.873 ± 0.005	0.829 ± 0.056
	Co-HITS	0.699 ± 0.047	0.733 ± 0.018	0.707 ± 0.062
	HITS	0.673 ± 0.003	0.520 ± 0.006	0.626 ± 0.043
	BGRM	0.559 ± 0.003	0.513 ± 0.001	0.614 ± 0.042
	BiRank	0.544 ± 0.003	0.504 ± 0.008	0.576 ± 0.031

Table 5 Account classification F1. Best results for each dataset are highlighted in darker shades

Sharing network	Method	Twitter_Covid	Twitter_Midterm	Facebook_Midterm
Reshare	node2vec	0.918 ± 0.003	0.886 ± 0.003	N/A
	LoCred	0.786 ± 0.005	0.757 ± 0.034	
	Rep. Scaling	0.576 ± 0.006	0.514 ± 0.007	
	Trustrank	0.196 ± 0.004	0.214 ± 0.002	
	PPR Trust	0.196 ± 0.004	0.218 ± 0.002	
	PR Trust	0.176 ± 0.012	0.251 ± 0.002	
Bipartite/Co-share	CoCred	0.800 ± 0.022	0.778 ± 0.014	0.756 ± 0.032
	node2vec	0.707 ± 0.024	0.886 ± 0.008	0.827 ± 0.029
	Co-HITS	0.638 ± 0.016	0.655 ± 0.019	0.728 ± 0.051
	HITS	0.738 ± 0.004	0.199 ± 0.010	0.628 ± 0.066
	BGRM	0.501 ± 0.002	0.146 ± 0.002	0.411 ± 0.121
	BiRank	0.477 ± 0.011	0.163 ± 0.089	0.434 ± 0.106

best-performing algorithms are LoCred and Reputation Scaling; these algorithms consider reshares as information diffusion channels without assuming good judgment by accounts. TrustRank, Personalized PageRank Trust, and PageRank Trust, which assume that reshares capture trust and hence reputation, perform only marginally better than random at the task. These results reinforce the difference between web-surfing behavior and news-sharing behavior on social media, where the assumption that good accounts seldom trust bad ones might not hold.

The bipartite/co-share networks capturing account trust in sources also provide effective information to estimate account credibility. The best-performing algorithm in each dataset achieves a ROC_AUC score of more than 0.8 (CoCred in the Twitter_Covid dataset and node2vec in the Twitter_Midterm and Facebook_Midterm datasets). Among the centrality-based algorithms on the bipartite network, our proposed algorithm CoCred performs the best, followed by Co-HITS and HITS; BGRM and BiRank perform close to random. CoCred's high ROC_AUC score confirms our intuition that the known account labels provide strong signals and should be preserved in score propagation to produce accurate ratings.

In summary, account credibility can be estimated accurately using trust signals from news-sharing networks. While all networks provide useful signals, the reshare network is slightly more informative for the task than the bipartite/co-share networks, based on the best-performing algorithm in each dataset.

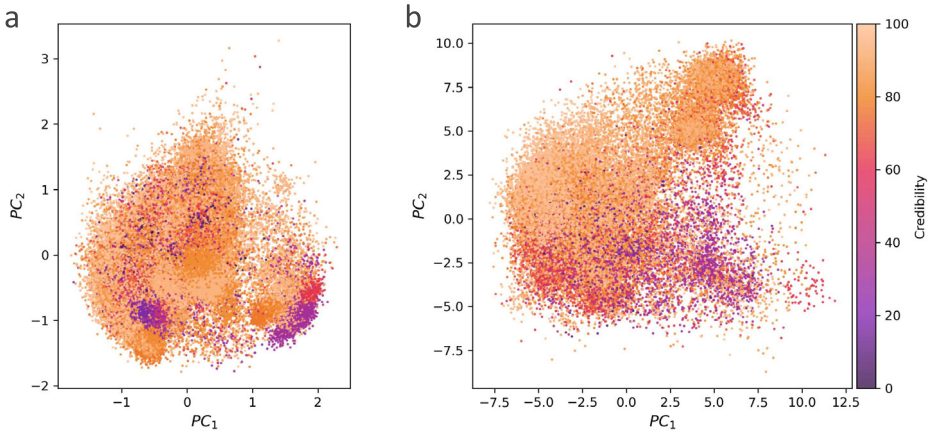

Figure 5 *Account embedding vectors for the Twitter_Covid dataset.* Node2vec vectors representing accounts are projected onto a two-dimensional space using PCA for (a) reshare and (b) co-share networks. Colors represent account credibility scores: orange and purple for high- and low-credibility accounts respectively

Figure 5 shows node2vec embedding vectors of accounts, with their dimensionality reduced using Principle Component Analysis (PCA) [60]. We observe that the embeddings in the reshare network more effectively cluster accounts with similar credibility, compared to the embeddings in the co-share network.

5 Discussion

This paper suggests ways to mitigate misinformation on social media by detecting low-credibility accounts likely to spread misinformation. To this end, we introduce methods to infer account credibility based on two information-sharing networks: a reshare network capturing account trust in other accounts and a bipartite network reflecting account trust in information sources. A systematic evaluation shows that the proposed methods are effective in detecting low-credibility accounts. The most successful method utilizes node2vec embeddings of accounts on reshare networks, with both ROC_AUC and F1 around 0.9. Our proposed network centrality algorithms, LoCred and CoCred, consistently perform well across the three datasets considered, with ROC_AUC and F1 both around 0.8. We find two kinds of credibility homophily in the news-sharing networks, which help explain the effectiveness of these methods: unreliable accounts tend to reshare content from one another (**Q1**), and share content from similar sources (**Q2**). These results confirm our hypothesis that the structure of the reshare or co-share network provides strong signals to effectively infer account credibility (**Q3**). While the presence of credibility homophily has also been observed in prior analyses of different Twitter data [61], future work might consider whether these results generalize to information-sharing networks derived from other social media platforms.

While our framework detects potential misinformation spreaders regardless of intention, it can help to better characterize disinformation spreaders in combination with other effective methods, such as bot and inauthentic coordinated campaign detection [62, 63]. In particular, one of our proposed methods, CoCred, provides source rankings simultaneously with account rankings. These source rankings can be used to estimate the credibility of emerging news media outlets. Such source credibility scores can be used as a feature

in misinformation classification, complementing extant content-based methods to curb misinformation.

Our work has some limitations. First, we only label and investigate accounts sharing sources with ratings by NewsGuard. Additional sources of misinformation that are not labeled by NewsGuard, as well as false/misleading claims in individual posts that do not include a link to some source, are missed in our analysis. However, since the criteria used by NewsGuard to label sources are independent from the methods we evaluate, there is no reason to believe that our accuracy measurements are biased. Still, complementary analyses using other sources of ground truth, such as post-level or account-level annotations, could provide additional insights. Second, classification results depend on the news-sharing data from which a network is derived. As such, account credibility scores may vary across contexts. For example, users have a higher tendency to share links to unconfirmed sources [64] when information is scarce, diverging from their usual news-sharing behavior. Accounts might be classified as unreliable in these situations, despite being reliable in others. One way to address this concern might be to aggregate an account's credibility scores across diffusion networks from different topics. Lastly, algorithms using the reshare network cannot be evaluated on the Facebook platform, as the data does not provide information about reshares. Furthermore, Twitter/X has recently removed free data access for researchers, making this kind of analysis difficult to replicate in the future. Note, however, that the proposed methods are platform-agnostic. The outlined analyses can be applied to data from any platform that allows for the construction of reshare or bipartite networks, such as Mastodon and Bluesky. We urge all social media platforms to make data available to allow for transparent analyses [65].

Overall, this work enriches the understanding of misinformation spreaders on social media by investigating trust signals in different information-sharing networks. The observed credibility homophily invites further exploration of misinformation diffusion and spreader dynamics. The proposed methods can be used to identify accounts of interest based on their estimated credibility by considering only sharing metadata, before further scrutiny (or even without any scrutiny) of the specific content they share. Platforms can readily apply our proposed methods to enhance their moderation efforts. These insights are crucial to social media platforms and policymakers in the debate on ways to combat online misinformation.

Acknowledgements
We are grateful to Kai-Cheng Yang for help with the data collection and many helpful comments. Ruj Akavipat and Pik-Mai Hui provided helpful discussions. We also thank NewsGuard for licensing their source credibility scores.

Funding
This research is supported in part by the Knight Foundation, Craig Newmark Philanthropies, DARPA (awards W911NF-17-C-0094 and HR001121C0169), and the Luddy School of Informatics, Computing, and Engineering at Indiana University, Bloomington.

Abbreviations
BGRM, Bipartite Graph Reinforcement Model; CoCred, Co-sharing network-based Credibility; HITS, Hyperlink-Induced Topic Search; KNN, k-Nearest Neighbors; LoCred, Low Credibility Account Estimation; P2P, peer-to-peer; PCA, Principle Component Analysis; PPR Trust, Personalized PageRank Trust; PR Trust, PageRank Trust; ROC_AUC, receiver operating characteristic curve; TF-IDF, Term Frequency — Inverse Document Frequency.

Data availability
For reproducibility, we make the data and code available in a public repository https://github.com/osome-iu/credibility-inference.

Declarations

Ethics approval and consent to participate
This research is based on analyses of public social media data with minimal risks to participants. This study has been granted exemption from review by the Indiana University IRB (protocol 17036). The collection and release of the dataset are in compliance with the platforms' terms of service.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author contributions
FM and BT formulated the research, developed the methodology, and prepared the manuscript. BT collected the data. BT and OA performed analyses. All authors read and approved the final manuscript.

Author details
[1]Observatory on Social Media, Indiana University, Bloomington, USA. [2]Network Science Institute, Northeastern University, Boston, USA.

Received: 11 September 2023 Accepted: 22 January 2024 Published online: 31 January 2024

References

1. Gottfried J, Shearer E (2016) News use across social media platforms 2016. pewresearch.org/journalism/2016/05/26/news-use-across-social-media-platforms-2016/
2. Zarocostas J (2020) How to fight an infodemic. Lancet 395(10225):676
3. Woolley SC, Howard PN (2018) Computational propaganda: political parties, politicians, and political manipulation on social media. Oxford University Press, London
4. Fisher M (2013) Syrian hackers claim AP hack that tipped stock market by $136 billion. Is it terrorism. washingtonpost.com/news/worldviews/wp/2013/04/23/syrian-hackers-claim-ap-hack-that-tipped-stock-market-by-136-billion-is-it-terrorism/
5. Tasnim S, Hossain MM, Mazumder H (2020) Impact of rumors and misinformation on COVID-19 in social media. J Prev Med Public Health 53(3):171–174
6. Allington D, Duffy B, Wessely S, Dhavan N, Rubin J (2021) Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. Psychol Med 51(10):1763–1769. https://doi.org/10.1017/S003329172000224X
7. Pierri F, Perry BL, DeVerna MR, Yang K-C, Flammini A, Menczer F, Bryden J (2022) Online misinformation is linked to early COVID-19 vaccination hesitancy and refusal. Sci Rep 12(1):5966
8. Yang K-C, Pierri F, Hui P-M, Axelrod D, Torres-Lugo C, Bryden J, Menczer F (2021) The COVID-19 infodemic: Twitter versus Facebook. Big Data Soc 8(1):20539517211013861
9. Zhou X, Jain A, Phoha VV, Zafarani R (2020) Fake news early detection: a theory-driven model. Digit Treats Res Pract 1(2):1–25
10. Goldstein JA, Sastry G, Musser M, DiResta R, Gentzel M, Sedova K (2023) Generative language models and automated influence operations: emerging threats and potential mitigations. arXiv preprint. arXiv:2301.04246
11. Menczer F, Crandall D, Ahn Y-Y, Kapadia A (2023) Addressing the harms of AI-generated inauthentic content. Nat Mach Intell. https://doi.org/10.1038/s42256-023-00690-w
12. Kirchner JH, Ahmad L, Aaronson S, Leike J (2023) New AI classifier for indicating AI-written text. OpenAI. openai.com/blog/new-ai-classifier-for-indicating-ai-written-text/
13. Hovland CI, Weiss W (1951) The influence of source credibility on communication effectiveness. Public Opin Q 15(4):635–650
14. Westerman D, Spence PR, Van Der Heide B (2014) Social media as information source: recency of updates and credibility of information. J Comput-Mediat Commun 19(2):171–183
15. Turcotte J, York C, Irving J, Scholl RM, Pingree RJ (2015) News recommendations from social media opinion leaders: effects on media trust and information seeking. J Comput-Mediat Commun 20(5):520–535
16. Yaqub W, Kakhidze O, Brockman ML, Memon N, Patil S (2020) Effects of credibility indicators on social media news sharing intent. In: Proc. 2020 CHI conf. on human factors in computing systems, pp 1–14. https://doi.org/10.1145/3313831.3376213
17. The Media Insight Project (2017) "Who Shared It?": how Americans decide what news to trust on social media. apnorc.org/projects/who-shared-it-how-americans-decide-what-news-to-trust-on-social-media/
18. Sterrett D, Malato D, Benz J, Kantor L, Tompson T, Rosenstiel T, Sonderman J, Loker K (2019) Who shared it?: deciding what news to trust on social media. Dig Journal 7(6):783–801
19. Mishra A, Bhattacharya A (2011) Finding the bias and prestige of nodes in networks based on trust scores. In: Proc. 20th intl. conf. on World Wide Web (WWW), pp 567–576. https://doi.org/10.1145/1963405.1963485
20. Rath B, Gao W, Ma J, Srivastava J (2018) Utilizing computational trust to identify rumor spreaders on Twitter. Soc Netw Anal Min 8(1):1–16
21. Bild DR, Liu Y, Dick RP, Mao ZM, Wallach DS (2015) Aggregate characterization of user behavior in Twitter and analysis of the retweet graph. ACM Trans Internet Technol 15(1):1–24
22. Shu K, Bernard HR, Liu H (2019) Studying fake news via network analysis: detection and mitigation. In: Emerging research challenges and opportunities in computational social network analysis and mining, pp 43–65
23. Shu K, Wang S, Liu H (2019) Beyond news contents: the role of social context for fake news detection. In: Proceedings of the twelfth ACM international conference on web search and data mining, pp 312–320

24. Al-Sharawneh J, Sinnappan S, Williams M-A (2013) Credibility-based Twitter social network analysis. In: Proc. Asia-Pacific web conf., pp 323–331
25. Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazer D (2019) Fake news on Twitter during the 2016 U.S. presidential election. Science 363(6425):374–378. https://doi.org/10.1126/science.aau2706
26. DeVerna MR, Aiyappa R, Pacheco D, Bryden J, Menczer F (2022) Identification and characterization of misinformation superspreaders on social media. Preprint. arXiv:2207.09524. https://doi.org/10.48550/ARXIV.2207.09524
27. Page L, Brin S, Motwani R, Winograd T (1999) The pagerank citation ranking: bringing order to the web. Technical report, Stanford InfoLab
28. Haveliwala TH (2003) Topic-sensitive pagerank: a context-sensitive ranking algorithm for web search. IEEE Trans Knowl Data Eng 15(4):784–796
29. Kamvar SD, Schlosser MT, Garcia-Molina H (2003) The eigentrust algorithm for reputation management in p2p networks. In: Proc. 12th intl. conf. on World Wide Web (WWW), pp 640–651. https://doi.org/10.1145/775152.775242
30. Zhou R, Hwang K (2007) Powertrust: a robust and scalable reputation system for trusted peer-to-peer computing. IEEE Trans Parallel Distrib Syst 18(4):460–473. https://doi.org/10.1109/TPDS.2007.1021
31. Gyongyi Z, Garcia-Molina H, Pedersen J (2004) Combating web spam with trustrank. In: Proc. 30th intl. conf. on very large data bases (VLDB). http://ilpubs.stanford.edu:8090/770/
32. Wang G, Wu J (2011) Flowtrust: trust inference with network flows. Front Comput Sci 5(2):181. https://doi.org/10.1007/s11704-011-0323-4
33. Akavipat R (2009) Distrust reputation system for P2P information sharing. PhD thesis, Indiana University. UMI Number: 3390252. https://proxyiub.uits.iu.edu/login?qurl=https
34. Ortega FJ, Troyano JA, Cruz FL, Vallejo CG, Enríquez F (2012) Propagation of trust and distrust for the detection of trolls in a social network. Comput Netw 56(12):2884–2895. https://doi.org/10.1016/j.comnet.2012.05.002
35. Guha R, Kumar R, Raghavan P, Tomkins A (2004) Propagation of trust and distrust. In: Proc. 13th intl. conf. on World Wide Web (WWW), pp 403–412. https://doi.org/10.1145/988672.988727
36. Kleinberg JM (1999) Authoritative sources in a hyperlinked environment. J ACM 46(5):604–632
37. Deng H, Lyu MR, King I (2009) A generalized co-hits algorithm and its application to bipartite graphs. In: Proc. 15th ACM SIGKDD intl. conf. knowledge discovery and data mining, pp 239–248
38. Rui X, Li M, Li Z, Ma W-Y, Yu N (2007) Bipartite graph reinforcement model for web image annotation. In: Proc. 15th ACM intl. conf. on multimedia, pp 585–594
39. He X, Gao M, Kan M-Y, Wang D (2016) Birank: towards ranking on bipartite graphs. IEEE Trans Knowl Data Eng 29(1):57–71
40. Romero DM, Galuba W, Asur S, Huberman BA (2011) Influence and passivity in social media. In: Proc. joint European conf. on machine learning and knowledge discovery in databases (ECML PKDD), pp 18–33. https://doi.org/10.1007/978-3-642-23808-6_2
41. Agichtein E, Castillo C, Donato D, Gionis A, Mishne G (2008) Finding high-quality content in social media. In: Proc. 2008 intl. conf. on web search and data mining, pp 183–194
42. Castillo C, Mendoza M, Poblete B (2013) Predicting information credibility in time-sensitive social media. Internet Res 23(5):560–588. https://doi.org/10.1108/IntR-05-2012-0095
43. Gupta A, Kumaraguru P, Castillo C, Meier P (2014) Tweetcred: real-time credibility assessment of content on Twitter. In: Proc. intl. conf. on social informatics, pp 228–243
44. Setiawan EB, Widyantoro DH, Surendro K (2020) Measuring information credibility in social media using combination of user profile and message content dimensions. Int J Comput Electr Eng 10(4):3537–3549. https://doi.org/10.11591/ijece.v10i4.pp3537-3549
45. Barbier G, Liu H (2011) Information provenance in social media. In: Proc. intl. conf. on social computing, behavioral-cultural modeling, and prediction, pp 276–283
46. Grover A, Leskovec J (2016) Node2vec: scalable feature learning for networks. In: Proc. 22nd ACM SIGKDD intl. conf. knowledge discovery and data mining, pp 855–864. https://doi.org/10.1145/2939672.2939754
47. Perozzi B, Al-Rfou R, Skiena S (2014) Deepwalk: online learning of social representations. In: Proc. 20th ACM SIGKDD intl. conf. knowledge discovery and data mining, pp 701–710. https://doi.org/10.1145/2623330.2623732
48. Tang J, Qu M, Wang M, Zhang M, Yan J, Mei Q (2015) Line: large-scale information network embedding. In: Proc. 24th intl. conf. on World Wide Web, pp 1067–1077
49. Hamdi T, Slimi H, Bounhas I, Slimani Y (2020) A hybrid approach for fake news detection in Twitter based on user features and graph embedding. In: Proc. intl. conf. on distr. comp. and Internet technology, pp 266–280
50. Roy A, Sarkar C, Srivastava J, Huh J (2016) Trustingness & trustworthiness: a pair of complementary trust measures in a social network. In: Proc. ACM/IEEE intl. conf. on advances in social networks analysis and mining (ASONAM), pp 549–554. https://doi.org/10.1109/ASONAM.2016.7752289
51. Zhao L, Hua T, Lu C-T, Chen I-R (2016) A topic-focused trust model for Twitter. Comput Commun 76:1–11. https://doi.org/10.1016/j.comcom.2015.08.001
52. Adali S, Escriva R, Goldberg M, Hayvanovych M, Magdon-Ismail M, Szymanski B, Wallace W, Williams G (2010) Measuring behavioral trust in social networks. In: Proc. IEEE intl. conf. on intelligence and security informatics, pp 150–152. https://doi.org/10.1109/ISI.2010.5484757
53. Altman NS (1992) An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat 46(3):175–185
54. Jones KS (1972) A statistical interpretation of term specificity and its application in retrieval. J Doc 28(1):11–21
55. Yang K-C, Torres-Lugo C, Menczer F (2020) Prevalence of low-credibility information on Twitter during the COVID-19 outbreak. In: Proc. ICWSM intl. workshop on cyber social threats (CySoc). https://doi.org/10.36190/2020.16
56. Aiyappa R, DeVerna MR, Pote M, Truong BT, Zhao W, Axelrod D, Pessianzadeh A, Kachwala Z, Kim M, Seckin OC et al (2023) A multi-platform collection of social media posts about the 2022 us midterm elections. In: Proceedings of the international AAAI conference on web and social media, vol 17, pp 981–989
57. Newman ME (2003) Mixing patterns in networks. Phys Rev E 67(2):026126
58. Alvarez-Hamelin JI, Dall'Asta L, Barrat A, Vespignani A (2006) Large scale networks fingerprinting and visualization using the k-core decomposition. In: Advances in neural information processing systems, pp 41–50
59. Serrano MÁ, Boguná M, Vespignani A (2009) Extracting the multiscale backbone of complex weighted networks. Proc Natl Acad Sci 106(16):6483–6488

60. Labrín C, Urdinez F (2020) Principal component analysis. In: R for political data science, pp 375–393
61. Nikolov D, Flammini A, Menczer F (2021) Right and left, partisanship predicts (asymmetric) vulnerability to misinformation. HKS Misinform Rev 1(7). https://doi.org/10.37016/mr-2020-55
62. Yang K-C, Varol O, Davis CA, Ferrara E, Flammini A, Menczer F (2019) Arming the public with artificial intelligence to counter social bots. Hum Behav Emerg Technol 1(1):48–61
63. Pacheco D, Hui P-M, Torres-Lugo C, Truong BT, Flammini A, Menczer F (2021) Uncovering coordinated networks on social media: methods and case studies. In: Proc. intl. AAAI conf. on web and social media (ICWSM), vol 15, pp 455–466
64. Oh O, Kwon KH, Rao HR (2010) An exploration of social media in extreme events: rumor theory and Twitter during the Haiti earthquake 2010. In: ICIS
65. Pasquetto IV, Swire-Thompson B et al (2020) Tackling misinformation: what researchers could do with social media data. HKS Misinform Rev 1(8). https://doi.org/10.37016/mr-2020-49

Publisher's Note