



The University of Texas at Austin
Center for Identity

A Survey on Quantitative Modeling of Trust in Online Social Networks

Wenting Song

K. Suzanne Barber

UTCID Report #26-01

February 2026

A Survey on Quantitative Modeling of Trust in Online Social Networks

WENTING SONG, University of Texas at Austin, USA

K. SUZANNE BARBER, University of Texas at Austin, USA

Online social networks facilitate user engagement and information sharing but are also rife with misinformation and deception. Research on trust modeling in online social networks focuses on developing computational models or algorithms to measure trust relationships, assess the reliability of shared content, and detect spam or malicious activities. However, most existing review papers either briefly mention the concept of trust or focus on a single category of trust models. In this paper, we offer a comprehensive categorization and review of state-of-the-art trust models developed for online social networks. First, we explore theories and models related to trust in psychology and identify several factors that influence the formation and evolution of online trust. Next, state-of-the-art trust models are categorized based on their algorithmic foundations. For each category, the modeling mechanisms are investigated, and their unique contributions to quantitative trust modeling are highlighted. Subsequently, we provide an implementation-centric trust modeling handbook, which summarizes available datasets, trust-related features, promising modeling techniques, and feasible application scenarios. Finally, the findings of the literature review are summarized, and unresolved challenges are discussed.

CCS Concepts: • **General and reference** → **Surveys and overviews**; • **Human-centered computing** → **Social networks**; *Social media*; *Social network analysis*; • **Security and privacy** → *Trust frameworks*; Social network security and privacy.

Additional Key Words and Phrases: computational trust, trust models, disinformation, uncertainty, graph neural networks, information diffusion, group decision making, crowdsourcing

ACM Reference Format:

Wenting Song and K. Suzanne Barber. 2026. A Survey on Quantitative Modeling of Trust in Online Social Networks. 1, 1 (February 2026), 34 pages. <https://doi.org/XXXXXXX.XXXXXXX>

1 INTRODUCTION

Online Social Networks (OSNs) are digital platforms that facilitate the creation, sharing, and exchange of information, ideas, and content among users. Nowadays, people tend to rely on online social networks and online community applications for daily social interactions. OSNs simplify the process for people to participate in online social activities and eliminate restrictions regarding when and where content can be created and shared. Nevertheless, online social networks are also rife with disinformation, misinformation, and deception [187]. User-generated content in online social networks exhibits significant variation in quality due to a variety of factors. Online users come from diverse backgrounds and have varying levels of expertise in the topics they discuss. Some may provide accurate and insightful information, while others may lack expertise and unknowingly share misleading information. Additionally, users often have personal biases and subjective opinions,

Authors' Contact Information: Wenting Song, University of Texas at Austin, Austin, Texas, USA, wentingsong@utexas.edu; K. Suzanne Barber, University of Texas at Austin, Austin, Texas, USA, sbarber@identity.utexas.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2026 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM XXXX-XXXX/2026/2-ART

<https://doi.org/XXXXXXX.XXXXXXX>

which may affect the quality of their contributions. The motivation behind creating content also affects its quality. Some users may intentionally create low-quality or misleading content for reasons such as trolling, spreading propaganda, or creating clickbait. In recent years, the emergence of generative AI models such as Generative Adversarial Networks (GANs) [48] or Variational Autoencoders (VAEs) [78] has led to a surge in online content. With the massive amount and diversity of user-generated and AI-generated content, it is becoming increasingly difficult, if not impossible, for users to distinguish authenticity and find valuable information. With all these issues in mind, it is imperative to critically evaluate online content by examining its source, accuracy, and relevance.

Trust is the bedrock of any relationship or connection, whether online or offline. Online trust plays a vital role in developing meaningful connections in online social networks. When trust is established, users are more likely to connect with others and cultivate reliable and productive interactions. The rise of trustworthy AI research emphasizes several trust-related factors, including robustness, generalization, explainability, transparency, repeatability, fairness, accountability, privacy, and security [77, 88]. Current research on trustworthy AI primarily focuses on establishing guidelines and standards for AI technologies [99]. In contrast, trust modeling in online social networks focuses on developing computational models and algorithms to evaluate trust relationships, verify the credibility of information, detect malicious or spam activities, and cultivate more trustworthy and high-quality online communities.

Research on trust modeling in online social networks is booming, but a systematic review of state-of-the-art trust models is still lacking. We are working to address this gap. In this survey paper, we will review the latest trust models developed for online social networks and explore the following questions:

- **RQ1:** What is trust? How is trust defined in online social networks?
- **RQ2:** What information or data is available to compute online trustworthiness?
- **RQ3:** How is online trust quantified?
- **RQ4:** What application scenarios benefit from trust quantification models?

Figure 1 provides an overview of this survey paper. First, we unpack the concept of trust and examine relevant psychological theories and models. Drawing on these theories, we identify key factors that influence the formation and evolution of trust, laying the foundation for trust modeling. Next, we examine trust models from different application domains to gain a more comprehensive understanding of the role of trust in online social networks. The state-of-the-art trust models are categorized into ten representative algorithmic categories. Throughout the literature review, we emphasize the practical implementation of trust models, focusing on data prerequisites, algorithmic foundations, and contextualized application scenarios. Thereafter, we provide an implementation-centric trust modeling handbook to inspire the development and implementation of trust models.

In summary, the **main contributions** of this paper are as follow:

- The concept of trust is unpacked, and key factors influencing the formation and evolution of online trust are identified.
- A comprehensive categorization and review of state-of-the-art research on trust modeling in online social networks.
- An implementation-centric trust modeling handbook summarizing available datasets, trust-related features, promising modeling techniques, and feasible application scenarios.

The paper is organized as follows. Section 2 describes related work in terms of other surveys reviewing trust modeling. The definition and influencing factors of online trust are explored in Section 3. In Section 4, we review the development and evolution of trust models over the past two

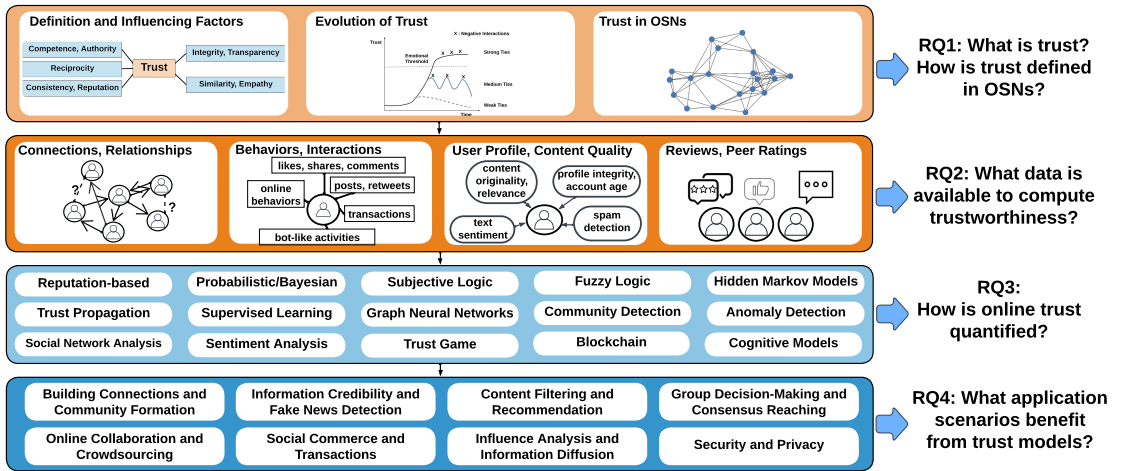


Fig. 1. The overall structure of this survey paper.

decades and discuss their trends. Section 5 provides a literature review of several representative state-of-the-art trust models developed for online social networks. The summary table in Section 6 outlines the data prerequisites, algorithmic foundations, and application scenarios of these trust models. In Section 7, we provide an implementation-centric trust modeling handbook, which summarizes available datasets, trust-related features, promising modeling techniques, and feasible application scenarios. The findings are discussed in Section 8. Lastly, Section 9 highlights the open challenges and research directions and Section 10 concludes the paper.

2 RELATED WORK

There are literature reviews on trust modeling in various fields, including agent systems [132], cybersecurity [56], the Internet of Things [5, 41], and online behavior research [54, 186]. Krausman et al. [81] reviewed trust measurement principles and proposed a methodological toolkit to facilitate the development of new trust models for human-autonomy teams [64]. Fan et al. [38] studied trust aggregation models and analyzed the resilience of trust metrics in the face of common threats in decentralized trust management systems. This review focuses on trust modeling in online social networks. First, we unpack the concept of trust and explore theories and models related to trust in psychology. Drawing on these theories, we identify several factors that influence the formation and evolution of online trust, laying the foundation for trust modeling. In this review, we focus on computational models that shed light on how trust is formed and evolves in online communities.

According to our study, a comprehensive literature review of the latest trust models in online social networks is still lacking. Some were published several years ago and may not capture the latest advances [23, 139]. Some delve into specific categories of trust models [16, 44, 69, 156], lacking a holistic perspective. This survey paper aims to fill this gap. We conduct a systematic review of the state-of-the-art trust models developed for online social networks and categorize these models into ten categories based on their algorithmic foundations. Furthermore, we observe that existing trust survey papers tend to propose the concept of trust but fail to elaborate on how to implement a trust model or how to apply the trust concept in practice [186]. In response, this review places particular emphasis on the development and implementation of trust models. In addition, we provide an implementation-centric trust modeling handbook that summarizes available datasets, trust-related features, promising modeling techniques, and feasible application scenarios.

3 DEFINITION AND FACTORS OF TRUST

This section focuses on the question: What is trust? How is trust defined in online social networks? Given that trust is essentially a psychological state, we first examine relevant theories and models in psychology. These theories offer valuable insights when modeling the establishment and development of trust within online communities. Next, we decompose the concept of trust and identify several factors that influence the evolution of online trust, laying the foundation for trust modeling.

3.1 What is Trust?

Julian Rotter defined trust as a general expectation that the words, promises, or statements of others can be relied upon [131]. Mayer et al. [110] define trustworthiness as a function of competence, benevolence, and integrity. As a psychological state, trust is a willingness to accept vulnerability based on positive expectations of another’s behavior [110]. Moreover, Blau’s social exchange theory [8] argues that trust develops through repeated social interactions with the expectation of mutual benefit. Lewicki and Bunker [87] proposed three stages of trust evolution over time: calculus-based trust, knowledge-based trust, and identification-based trust. Trust often begins with calculus-based trust, where individuals weigh potential risks against expected benefits. Through repeated interactions and positive experiences, trust evolves into knowledge-based trust. Ultimately, emotional connections and shared values foster identification-based trust, which can be seen as deeper relationships and stronger community bonds [111].

3.2 Trust Influencing Factors

Drawing on psychological theories and models related to trust, we highlight several factors that influence the formation and evolution of trust. As shown in Figure 2, the diversity of trust factors also highlights the multifaceted nature of trust.

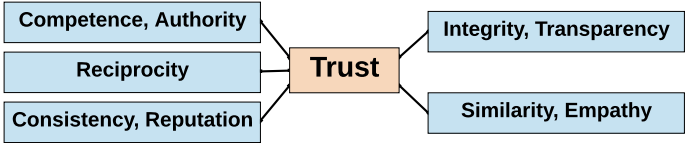


Fig. 2. Trust Influencing Factors

- **Competence, Authority**
Competence is emphasized as a key factor in various theories of trust [30, 110, 112]. Trust is established when a trustee demonstrates a high level of competence in achieving desired outcomes or is recognized as an authority in a field. If trustors perceive the trustee as knowledgeable or efficient, they will quickly develop trust [112].
- **Reciprocity**
According to Deutsch’s theory [30], trust is a decision made under uncertain conditions with potential risks and benefits. Trustors tend to assess the risks involved in an interaction and determine whether the potential benefits outweigh the risks [107, 186]. Trust grows when both parties are fair and reciprocal (Blau’s social exchange theory [8]).
- **Consistency, Reputation**
Trustees are expected to demonstrate consistent, reliable behavior over time [131]. In online social networks, users’ reputation is usually determined by their past behavior and numerical

indicators such as reviews, ratings, or peer endorsements. Predictable behavior enhances the perception of trustworthiness.

- Integrity, Transparency

Integrity, transparency, and clear communication are essential to building lasting trust [110]. Trust is built when the other party is open about their intentions, motivations, and information. For example, health apps are more likely to earn users' trust if they openly inform users about how their data is used and demonstrate a commitment to transparency and ethical practices.

- Similarity, Empathy

Trust stems from shared experiences, emotional connections, and empathy [111]. This emotional bond and shared values foster identification-based trust [87], which builds deep and lasting relationships.

3.3 Characteristics of Online Trust

In face-to-face interactions, trust is built through shared experiences and social cues such as eye contact and body language. In the absence of physical cues, online trust is primarily established through digital cues, including but not limited to reputation indicators (e.g., reviews, ratings, and user endorsements) and profile indicators (e.g., photos, bios, affiliations, and verified credentials). This section discusses the characteristics of online trust.

(1) Context Dependency and Domain Specificity

Trust varies by platform, community norms, and type of interaction [49]. A user might trust someone on LinkedIn for professional advice, but not someone on Instagram for financial advice.

(2) Trust Maintenance or Decay

Trust is built gradually through repeated positive interactions [87, 131]. Meanwhile, trust can decay quickly due to negative interactions or violations (e.g., deception, manipulation) [142] (Figure 3).

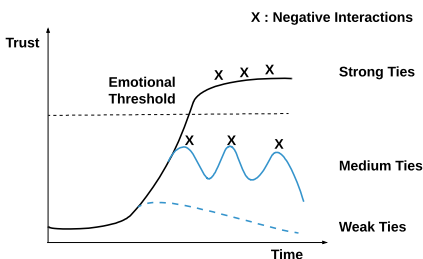


Fig. 3. **Development and Decay of Trust Over Time** [142]. Three types of ties are illustrated. Strong ties mean that trust is built through multiple interactions in a relatively short period of time, breaks through emotional thresholds, and can withstand negative interactions. Moderate ties mean that trust develops through fewer interactions, has not yet reached an emotional threshold, and will weaken if interactions cease or negative interactions persist. In weak ties, negative interactions or lack of interaction can lead to the end of the relationship.

(3) Anonymity and Iduser Cues

Online users often participate in a partially or fully anonymous manner, which makes building trust more difficult. While real names can enhance trust, anonymity may encourage more open and honest expression. Additionally, group affiliations or verification badges can serve as supplementary indicators of trustworthiness.

(4) Trust Propagation

Trust is transitive and can be transferred across connections, albeit with some reduction. Trust propagation is a fundamental component of e-commerce systems (e.g., eBay), aiming to help users find trustworthy reviewers, sellers, and products.

(5) Cognition and Affectiveness

McAllister [111] argues that trust includes both rational evaluation (cognition-based trust) and emotional bonds (affect-based trust). Cognitive trust stems from rational judgment and observable evidence, such as evaluating someone's competence or contributions. Affective trust, on the other hand, stems from emotional connections, including empathy, intimacy, and shared values [87].

(6) Interpersonal Trust and System-based Trust

Interpersonal trust refers to trust in a specific individual. In contrast, system-based trust [47], also known as institutional trust, refers to the trust that individuals place in institutions (e.g., online platforms). It encompasses users' trust in the platform's ability to manage moderation, protect privacy, and regulate content (e.g., flagging or removing harmful content).

4 EVOLUTION OF TRUST MODELS

Before delving into specific trust models, we will first review the evolution of trust models over the past two decades. Figure 4 provides an overview of the evolution of trust models developed for online social networks since the 2000s.

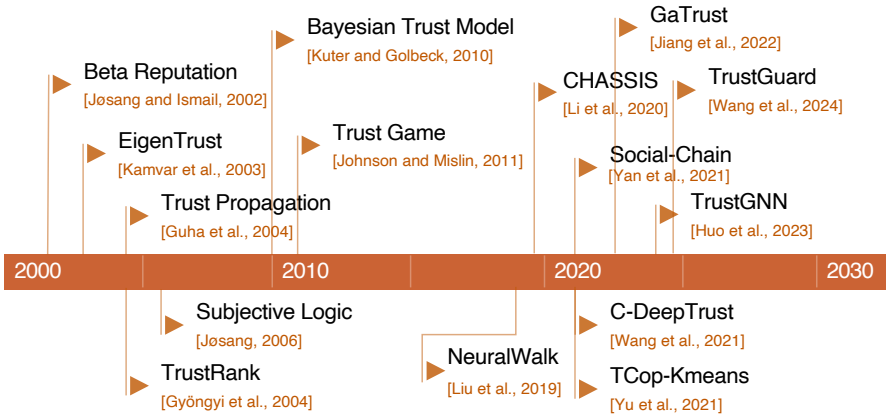


Fig. 4. The evolution of trust models developed for online social networks from the 2000s to the 2020s.

Initially, trust modeling was an interdisciplinary study that combined psychology and human factors [110, 131]. In the 2000s, research on trust modeling in online social networks drew inspiration from the conceptual definition of trust and focused on modeling the basic attributes of trust. Probabilistic and subjective logic trust models have been proposed to model the subjectivity and uncertainty in trust relationships [22, 72, 73]. In reputation trust models [55, 75], users accumulate trust scores through their behavior, contributions, and peer feedback during interactions. Various trust models have collectively enhanced our understanding of how online trust is established, maintained, or destroyed.

Over the past decade, the field has gradually shifted towards data-driven and machine learning approaches. With the development of deep learning models and the continuous emergence of large-scale datasets, many trust models have adopted techniques such as feature-based supervised learning models [156], Graph neural networks [68, 93], K-means clustering [176], and anomaly detection [163] to quantify trust. In recent years, the techniques and application scenarios of trust models have become increasingly diversified. Decentralized trust evaluation models based on blockchain [38, 150], cognitive trust models [90], trust models developed for misinformation detection [71, 141], and trust models for human-AI collaboration [148] have also emerged.

5 A SYSTEMATIC REVIEW OF TRUST QUANTIFICATION MODELS

This section systematically reviews the state-of-the-art research on trust modeling and categorizes the models into ten categories based on their algorithmic foundations. For each category, we investigate the underlying modeling mechanisms and highlight their unique contributions to the quantitative modeling of trust in online social networks.

5.1 Reputation-Based Trust Models

Early research on trust models focused on reputation-based models (e.g., EigenTrust [75] and TrustRank [55]). User trustworthiness is measured as a reputation score, which is accumulated through behavior, contributions, and peer feedback during interactions. Reputation-based trust models aggregate community opinions, and the computation process is transparent to users. However, such models are vulnerable to Sybil attacks [118], collusion, and fake reviews.

Rezvani and Rezvani [128] proposed a novel reputation system to distinguish fake and trustworthy consumer ratings. Inspired by the iterative filtering algorithm [29], users whose ratings frequently deviate from those of other users will be assigned lower weights. The proposed random iterative filtering algorithm simultaneously calculates user weights and aggregates reputation scores. Using only a random subset of ratings, it can approximate the average rating for each product within a certain error threshold. This randomness also makes the system resistant to unfair ratings and simple attacks.

GOLI [65] aims to identify opinion leaders in online social networks by evaluating user reputation. Initially, it constructs a trust network by analyzing user characteristics and social connections. Subsequently, the reputation score is derived from feedback during interactions and trust ratings provided by neighbors. Finally, the aggregate reputation score is combined with in-degree centrality to identify influential opinion leaders.

Wu et al. [164] developed an optimal feedback model to prevent manipulation during group decision-making. Based on trust connections, each expert is assigned a behavioral weight, which determines their influence in reaching consensus. In addition, the model calculates optimal opinion adjustment costs to promote consistent behavior.

5.2 Probabilistic and Bayesian Trust Models

Probabilistic trust models represent trust as a probability distribution rather than a fixed number [83]. They handle the uncertainty of trust and have a solid mathematical foundation. For instance, Bayesian trust models use Bayesian inference to update the probability of a user being trustworthy based on new evidence or interactions [14, 25, 161]. Probabilistic trust models work well with sparse interactions or noisy data. However, they require prior distributions and can be computationally expensive.

Wang et al. [160] are committed to curbing the spread of negative information through intervention strategies. They defined four user states and calculated the state transition probabilities based on trust relationships. Three collaborative intervention strategies—warning, correction, and

guidance—were introduced to address users in different states: the unknown state, the negative information diffusion state, and the dual information hesitation state. Moreover, Naumzik and Feuerriegel [117] employed stochastic processes [43, 127] and Bayesian inference to model information diffusion and detect misinformation on social media. Inspired by the Hawkes process [57, 185], they developed a new probabilistic mixture model that can simulate retweet cascades and directly classify them as true or false, eliminating the need for feature engineering.

Lingam et al. [94] worked on social bots and botnet community detection on Twitter (X). First, a weighted graph was constructed, where nodes represent accounts and edges represent behavioral similarities and trust scores between accounts. Trust scores were calculated using Bayesian inference [25] and random walk stochastic processes [66]. Next, the authors leveraged graph theory and deep autoencoders to identify communities (groups of accounts) engaging in malicious behavior. Research also shows that bots tend to grow followers more slowly than real accounts and have a disproportionate follower-to-following ratio.

5.3 Subjectivity and Uncertainty-Based Trust Models

Subjective logic is a probabilistic logic that deals with the uncertainty in trust relationships [22, 72, 73]. Trust is represented by an opinion triple consisting of belief (b), disbelief (d), and uncertainty (u). Fuzzy logic deals with the ambiguity in trust using linguistic terms (e.g., "high trust" or "low trust") and membership functions [177, 178]. They are useful in situations where trust is not easily quantifiable. However, the definitions of opinion triples and membership functions may vary in different contexts and contain biases [58, 121].

Cho et al. [24] aims to reduce the spread of false information. They proposed an enhanced opinion model based on subjective logic to classify user opinions on false information into trust, distrust, and uncertainty. The simulation experiments revealed several factors that influence the spread of false information, including the ratio of information propagation frequency, centrality measures, opinion decay rate, and the proportion of true informants.

Ma et al. [105] proposed a novel trust-based consensus-reaching model for group decision-making. Intuitive fuzzy sets deal with uncertainty and ambiguity in the interaction between experts. The model assigned weights to experts using a second-order additive fuzzy measure combined with the Shapley value. Experts with higher weights had more impact on reaching a consensus. One limitation is that the model assumes that social trust relations are static.

3VSL-AT [97, 98] was designed for trust computation in online social networks. The authors proposed a three-valued subjective logic model to capture uncertainty in trust and used Bayesian inference to calculate trust relationships. Leveraging depth-first search (DFS), the AssessTrust algorithm can calculate the trust relationship between any pair of users.

5.4 Context-Aware Trust Models

Trust is situational, not universal. A user might trust someone for movie recommendations but not for financial advice. Context-aware trust models calculate trust scores specific to domains, topics, or situations rather than taking a one-size-fits-all approach [158, 187]. It is useful for personalized recommendations, content curation, and contextual decision support. The challenges come from the need to detect multiple contexts, increased data requirements, and more complex modeling.

DDTrust [45] was proposed as a dynamic trust relationship prediction model between user pairs. The authors employed two text analysis tools (LIWC [15, 124] and BERT [32]) to identify user sentiment and generate word embeddings from tweet contexts and user profiles. It improves trust relationship prediction by combining user sentiment analysis and textual content inspection. Similarly, DCAT [46] predicts pairwise trust relationships based on user demographics and textual content. LIWC [15, 124] and word embedding techniques were used for contextual analysis. In

addition, Gatrust [68] integrates node attributes, contextual features, network topology, and trust interactions into a graph neural network approach to improve the accuracy of trust relationship prediction.

Xue et al. [171] proposed a novel spam review detection model based on trust evaluation. The model evaluated user trustworthiness by analyzing review content and examining the opinions expressed. Key evaluation factors include the inconsistency of opinions in user reviews and the extent to which these reviews deviate from the majority. The study found that user trustworthiness can be inferred from the opinions they express in online reviews.

5.5 Game-Theoretic Trust Models

The core idea of game-theoretic trust models is that trust emerges from rational choices in repeated interactions. They model trust as the outcome of strategic interactions and apply game theory principles to analyze the interactions between users [59, 70, 143]. Popular game-theoretic trust models include the prisoner's dilemma [82, 125, 126], Bayesian games, and signaling games [21]. It is well-suited for market-oriented social networks where incentives matter. Nonetheless, many platforms allow users to remain anonymous or create multiple idusers, reducing the effectiveness of repeated gaming strategies or penalties for breach of trust.

Mahajan et al. [106] proposed a novel trust-based team formation algorithm named PRADA-TF. It modeled the team formation process as a strategic game and adopted the Mechanism Design algorithm [12, 119]. PRADA-TF selected team members by evaluating their trustworthiness. The team formation process prioritizes domain expertise and user privacy preferences, aiming to optimize team performance while respecting individual privacy preferences.

To combat the spread of disinformation, Guo et al. [52, 53] introduced an opinion framework using repeated games. They modeled the interaction as a three-player game involving attackers, users, and defenders. The authors exploited subjective logic to model how users update their opinions under uncertainty and how social interactions influence the spread of disinformation. Similarly, Zhang et al. [180] proposed an evolutionary information diffusion model based on Evolutionary Game Theory (EGT) [4, 168], aiming to mitigate the adverse effects of malicious users.

5.6 Graph-Based Trust Models

Graph-based trust models describe social networks as weighted graphs, where users are represented as nodes and trust relationships are represented as weighted edges [34, 69, 108]. Trust is propagated or aggregated across network paths [51]. These models exploit network structure and trust transitivity to infer trust from connections and paths [170]. Typical use cases include friend recommendations and spammer detection. Graph-based trust models can effectively capture structural relationships, but require known explicit trust relationships to function.

The most common application of graph-based trust models is to predict trust relationships [93, 96, 159, 169]. Guardian [93] was introduced based on graph neural networks (GNNs), which learns the latent factors of trust propagation and trust aggregation. It predicts the trust relationship between any two users by examining known trust relationships and social network structures. To improve the accuracy of trust relationship prediction, AtNE-Trust [159] captures the characteristics of trust networks and multi-dimensional user attributes simultaneously. With the same goal, AHNTP [17] adopts hypergraphs to model high-order dependencies and integrates motif-based PageRank [184] to identify social influencers and enhance trust propagation.

Wu et al. [165] leveraged trust propagation models to calculate indirect trust relationships and construct a complete trust network. The calculated trust score serves as a weighting factor, simplifying and reducing the effort required to reach consensus in group decision-making.

5.7 Machine Learning and Data-Driven Trust Models

Machine learning trust models use features of interactions and network structure to learn and predict trustworthiness [156]. Supervised learning models rely on labeled trusted/untrusted data for training. Unsupervised learning models such as clustering and anomaly detection can identify unusual behavior patterns that may indicate untrustworthiness [163]. Graph neural networks (GNNs) learn trust embeddings directly from the social graph for trust prediction [68, 93]. Meanwhile, machine learning trust models are prone to bias, and some lack interpretability due to their complex structures.

PTP-MF [113] adopts matrix factorization to predict pairwise trust relationships. Each user is represented by a feature vector pair $\mathbf{P}_i$ and $\mathbf{Q}_i$. The trustor vector $\mathbf{P}_i$ represents the user's tendency to trust others. The trustee vector $\mathbf{Q}_i$ represents how others perceive the trustworthiness of the user. Experiments showed that PTP-MF can make accurate predictions even when the user connection graph is sparse.

Graph neural networks (GNNs) have emerged as a powerful approach for learning complex relationships and dependencies in graph-structured data [135, 153, 166, 167]. TrustGNN [63] is a representative GNN-based trust model that achieves satisfactory trust prediction accuracy and exhibits robustness against malicious users. The authors combined the graph attention mechanism [153] with trust propagation theory and proposed a trust-specific message passing where the attention weights reflect the trustworthiness. Trustguard [157] is another trust model that utilizes GNNs and a position-aware attention mechanism [183]. Differently, it captures trust dynamics and integrates defense mechanisms within the spatial aggregation layer, enhancing its ability to resist trust-related attacks.

TCop-Kmeans [176] is a Trust Cop-Kmeans clustering algorithm designed to support consensus reaching in large-scale decision making. It leveraged the similarity in opinions and trust relationships among decision-makers to optimize the clustering algorithm. Additionally, the trust relationship scores were used as weights to develop a minimum cost consensus model.

5.8 Blockchain and Decentralized Trust Models

Blockchain enables decentralized trust computation, immutability of interactions, and transparency in trust relationships [38, 56, 67, 150]. All transactions or interactions are recorded on an immutable, timestamped distributed ledger. Trust evaluation is conducted in a distributed manner on the blockchain, making the system more robust and reducing the risk of malicious tampering or Sybil attacks [133]. Based on blockchain technology, Sarode et al. [133] proposed a novel decentralized social network. The immutability of blockchain ensures that all transactions and content remain tamper-proof, protecting them from unauthorized modifications. In addition, the system incorporates incentive mechanisms to promote honest and reliable participation—users can be rewarded for creating valuable and accurate content, while malicious behavior will be punished.

Social-Chain [173] was proposed as a new blockchain-based decentralized trust evaluation model. The authors introduced a new consensus mechanism, Proof of Trust (PoT), where miners can generate new blocks by collecting new trust evidence. This trust evaluation system employed a distributed architecture, making it resistant to traditional attacks.

Yu et al. [175] proposed a distributed reputation mechanism to distinguish the authenticity of information in online social networks and prevent the spread of misinformation. User reputation was assessed based on the authenticity of the messages they sent. The authors used a Transformer-based language model (RoBERTa [102]) combined with sentiment analysis techniques to verify the authenticity of the messages. Blockchain smart contracts were utilized to automatically calculate and update user reputation scores.

5.9 Psychological and Cognitive Trust Models

McAllister [111] believes that building trust includes both rational evaluation (cognitive trust) and emotional bonds (affective trust). Cognitive trust models draw on psychology and cognitive science to simulate how users perceive, form, and update trust when interacting with other users or content in online social networks [19, 37, 39]. Granatyr et al. [50] argued that considering cognitive factors is crucial when constructing a trust model.

CHASSIS [90] applied the Hawkes process [57, 185] to simulate information diffusion. The authors incorporated conformity theory from social psychology into the information diffusion model [3, 26, 31] and quantified conformity by examining diffusion trees [89]. The introduced semi-parametric inference method helps identify which individuals exhibit conformity behavior.

Karami et al. [76] studied the characteristics and motivations of users who spread misinformation on social media. Drawing on psychological theories [33, 147], they identified five motivational factors: uncertainty, anxiety, lack of control, relationship enhancement, and social rank. The motivational factors help distinguish misinformation spreaders from ordinary users and enable the early detection and crackdown on the spread of misinformation.

5.10 Hybrid and Multi-Component Trust Models

The multi-component trust model assesses trust by comprehensively considering multiple factors. It leverages the advantages of multiple methods to achieve a more nuanced, robust, and accurate trust evaluation.

TrustLab™ [61, 62, 179] was introduced to measure individual trustworthiness on social media. Trust was assessed based on six trust attributes: authority, experience, expertise, iduser, proximity, and reputation. It calculated user trust scores from multiple dimensions, including user characteristics, historical posting behavior, domain expertise, physical proximity, and social network structure. Additionally, various charts were used to visualize the distribution of Twitter (X) user trust scores.

T-VLC [20] was proposed as a general trust model for assessing trust among participants in online learning communities. It measures trust across eight dimensions: direct experience, reputation, role, knowledge, security, quality, institutional trust, and closeness, with each dimension assigned an aggregated weight. As a flexible framework, the model allows these weights to be adjusted to suit different application scenarios.

iSim [42] was proposed as a multi-component trust prediction method based on user similarity. This method employed techniques such as vector space similarity, trust propagation, and time-aware matrix factorization for comprehensive trust assessment.

6 REPRESENTATIVE TRUST MODEL REVIEW

In this section, we summarize several state-of-the-art trust models (TM) for online social networks introduced between 2019 and 2025 in Table 1. Throughout the literature review, we emphasize the implementation aspects of trust modeling, focusing on data prerequisites, algorithmic foundations, and contextualized application scenarios.

Table 1. A Review of the State-of-the-Art Trust Models (TM) for Online Social Networks (2019-2025).

	Models	Data Requirements	Algorithms/Techniques	Application Scenarios
1. Reputation-Based TM	Rezvani and Rezvani [128] (2020)	user-provided ratings	iterative filtering [29]	to distinguish between trustworthy and fake consumer ratings
	GOLI [65] (2023)	user characteristics, number of followers, user connections, posts, response time	graph neural networks, in-degree centrality	to classify node importance and identify opinion leaders in online social networks
	Wu et al. [164] (2020)	existing trust relationships between users, user online behaviors	social network analysis, in-degree centrality	to prevent individual and group manipulation during the consensus-reaching process of group decision-making
2. Probabilistic/Bayesian TM	Wang et al. [160] (2020)	existing trust relationships between users	state transitions, Pontryagin's maximum principle	to suppress the spread of negative information through collaborative intervention strategies
	Naumzik and Feuerriegel [117] (2022)	retweet cascades, user tweets, number of followers	Hawkes process [57, 185], Bayesian inference [43, 127], Markov chain Monte Carlo sampling [43]	to detect false information on social media; classify information as true or false based on the propagation process
	DA-SBCD [94] (2020)	user social activities (tweets, retweets, posts), user profile, number of followers	Bayesian theorem [25], random walk stochastic process [66], deep autoencoders, similarity measures, sentiment analysis	to detect botnet communities and groups of accounts with malicious behavior
3. Subjectivity/Uncertainty TM	Cho et al. [24] (2019)	user connections within online social networks, user online behaviors	subjective logic, centrality measures, opinion decay	to mitigate false information and fake news on social networks
	Ma et al. [105] (2021)	interactive relationships between experts	Choquet integral, intuitionistic fuzzy set, Shapley value, centrality measures	to facilitate the consensus-reaching process in group decision-making
	3VSL-AT [97, 98] (2021)	user connections in online social networks	Bayesian inference, Dirichlet-categorical distribution, subjective logic (discounting and combining operators), depth-first search	to compute the trust relationship between any two users in an online social network
4. Context-Aware TM	Xue et al. [171] (2019)	user online reviews	SVM classifier, Word2Vec [115], three-layer trust propagation [154]	to detect spam reviews through trust evaluation; assess the trustworthiness of users, reviews, and statements by analyzing the opinions expressed in reviews
	DDTrust [45] (2020)	tweet context, user profile, number of followers	LIWC [15, 124], BERT [32], neural networks	to identify user emotions from tweet context; to predict trust relationships between user pairs over time
	Gatrust [68] (2022)	existing trust relationships between users, user contextual features	graph neural networks	to predict trust relationships between user pairs

continued on next page

	Models	Data Requirements	Algorithms/Techniques	Application Scenarios
5. Game-Theoretic TM	PRADA-TF [106] (2024)	user connections, user profile (domain expertise, privacy preferences)	Mechanism Design [12, 119], subjective logic (discounting and consensus operators), betweenness centrality	introduced a trust-oriented team formation algorithm that selects potential members by considering domain expertise and privacy preferences
	Guo et al. [52, 53] (2022)	user type (active user/fake bots/spammers), user interactions, tweets texts	Latent Dirichlet Allocation, subjective logic, Nash Equilibrium	proposed a game-theoretic opinion framework to study opinion dynamics and combat the spread of disinformation
	Zhang et al. [180] (2020)	user connections, user type (rational/malicious users)	evolutionary game theory, reputation mechanisms, social norm, death-birth updating rule	proposed an evolutionary model of information diffusion to mitigate the negative impact of malicious users
6. Graph-Based TM	Wu et al. [165] (2021)	existing trust relationships	social network analysis	construct a complete trust network using trust propagation models; calculated trust scores used as weights to promote consensus in group decision-making
	AHNTF [169] (2024)	existing trust relationships, user behaviors, user-provided ratings/reviews	hypergraphs [17], motif-based PageRank, cosine similarity [184]	to predict trust relationships; improve prediction accuracy by capturing high-order dependencies with hypergraphs
	AtNE-Trust [159] (2020)	existing trust relationships, user-provided ratings and reviews	skip-gram with negative sampling [115, 116], doc2vec [84], autoencoders, cosine similarity, matrix factorization	to predict pairwise trust relationships; combine trust network features and multi-dimensional user attributes to enhance trust relationship prediction
7. Machine Learning TM	PTP-MF [113] (2019)	existing trust relationships	matrix factorization, cognitive biases, gradient descent	to predict trust relationships between user pairs
	TrustGNN [63] (2023)	existing trust relationships	graph neural networks, attention mechanism [153]	to predict trust relationships; model various trust propagation patterns and construct different trust chains for a comprehensive trust evaluation
	TrustGuard [157] (2024)	temporal trust ratings from other users	graph neural networks, position-aware attention mechanism [183], trust propagation, similarity measures	to predict trust relationships; model trust dynamics; incorporate defense mechanisms to combat trust-related attacks
	TCop-Kmeans [176] (2021)	individual opinions, trust ratings from other users	social network analysis, clustering, similarity measures	leverage trust relationships as weights to develop a minimum-cost consensus model; facilitate consensus reaching in large-scale decision-making
8. Decentralized TM	Social-Chain [173] (2021)	social behaviors, communication interactions on the blockchain	public-private keys, consensus mechanisms, timestamp validation algorithms	proposed a robust blockchain-based decentralized trust evaluation model; introduced a new consensus mechanism named Proof-of-Trust (PoT)
	Yu et al. [175] (2023)	transactions and messages on the blockchain	smart contracts, sentiment analysis, RoBERTa [102], IPFS [7]	proposed a distributed reputation mechanism to distinguish the authenticity of information and prevent the spread of misinformation.

continued on next page

	Models	Data Requirements	Algorithms/Techniques	Application Scenarios
9. Cognitive TM	Sarode et al. [133] (2023)	n/a.	Ethereum blockchain, smart contracts, IPFS [7], SHA-256	introduced a novel decentralized social network; incentive mechanisms reward high-quality content contributions and promote reliable participation
	CHASSIS [90] (2020)	user online social activities (tweets, retweets, posts, comments, likes)	conformity [3, 26, 31], diffusion trees [89], Hawkes process [57, 185]	introduced a novel information diffusion model based on a social psychology theory named conformity; investigate the interplay between informational conformity and normative conformity
	Karami et al. [76] (2021)	user profile, number of followers/friends, user online social activities (tweets, retweets, posts)	rumor psychology [33], words psychology [147], LIWC features [15, 124], BERT [32]	to analyze the characteristics and motivations of individuals who spread false news on social media
	CPTT [39] (2020)	n/a.	predictive processing in cognitive theory [27, 155]	depict how individuals rely on cognitive predictions to form trust judgments during interactions and information exchanges
10. Multi-Component TM	TrustLab™ [61, 62, 179] (2021)	number of followers, user online social activities (tweets, retweets, posts), user characteristics	sentiment analysis, PageRank [120], HITS [79], Random Forests	measure user trustworthiness on social media with multiple trust attributes; visualize the distribution of Twitter (X) user trust scores
	iSim [42] (2021)	trust ratings of users to their peers	vector space similarity, time-aware matrix factorization, trust propagation, similarity measures	proposed a multi-component trust prediction method based on user similarity
	T-VLC [20] (2023)	participant characteristics, interactions between participants	Bayesian networks, social network analysis, centrality measures, Pearson coefficient	proposed a general trust model for assessing trust among participants in online learning communities

7 HANDBOOK FOR TRUST MODELING

As George Box stated, "all models are wrong, but some models are useful" [13]. In this section, we will provide an implementation-centric trust modeling handbook to inspire the development and implementation of trust models. Based on the literature review, we summarize the available datasets, trust-related features, promising modeling techniques, and feasible application scenarios. Overall, the effectiveness of trust modeling depends on the quality of the data, the relevance of the trust features, and the suitability of the chosen algorithm or approach.

7.1 Available Data to Compute Online Trustworthiness

In this section, we attempt to address the question: What information or data is available to compute online trustworthiness? We explored various variables and datasets from online social networks that may indicate online trustworthiness. Table 2 summarizes the datasets available for trust modeling and experimental evaluation. Relevant and diverse datasets provide the foundation for making reasonable assumptions in trust modeling. Selecting features that are highly correlated with trust and can effectively characterize trust is also crucial. In Figure 5, we outline four types of datasets commonly used to calculate online trust. For each type of dataset, we identified several trust-related features that can be used to build trust models.

Table 2. Available Datasets for Trust Modeling in Online Social Networks.

Datasets	Source Links	Details of the Datasets	User Profile	Trust Relationships	Connections	Interactions	Reviews, Ratings
Advogato [109]	https://networkrepository.com/advogato.php	6,541 users and 51,127 trust relationships. Four trust relationships {Observer, Apprentice, Journeyer, Master}		✓			
Pretty Good Privacy (PGP) [11]	https://networkrepository.com/tech-pgp.php	38,546 users and 317,979 trust relationships. Four levels of trust relationships.		✓			
BlogCatalog [146]	http://datasets.syr.edu/datasets/BlogCatalog3.html	10,312 bloggers (nodes), 39 groups and 333,983 friendship pairs (binary, undirected edges).	✓		✓		
Flickr [146]	http://datasets.syr.edu/datasets/Flickr.html	80,513 users (nodes), 195 groups and 5,899,882 friendship pairs (binary, undirected edges).	✓		✓		
Youtube [146]	http://datasets.syr.edu/datasets/YouTube2.html	1,138,499 users (nodes), 47 groups and 2,990,443 friendship pairs (binary, undirected edges).	✓		✓		
Social Circles: Facebook [86]	https://snap.stanford.edu/data/ego-Facebook.html	4,039 users (nodes) and 88,234 undirected edges.	✓		✓		
Social Circles: Twitter [86]	https://snap.stanford.edu/data/ego-Twitter.html	81,306 users (nodes) and 1,768,149 directed edges.	✓		✓		
Social Circles: Google+ [86]	https://snap.stanford.edu/data/ego-Gplus.html	107,614 users (nodes) and 13,673,453 directed edges.	✓		✓		
DBLP [145]	https://www.aminer.org/citation	5,259,858 papers and 36,630,661 citation relationships.	✓		✓		
Cora [136]	https://linqs.org/datasets/#cora	2,708 scientific publications, 7 classification classes and 5,429 links (binary, directed edges).	✓		✓		
Citeseer [130]	https://linqs.org/datasets/#citeseer-doc-classification	3,312 scientific publications, 6 classification classes and 4,732 links (binary, directed edges).	✓		✓		
PubMed	https://linqs.org/datasets/#pubmed-diabetes	19,717 scientific publications and 44,338 links (binary, directed edges).	✓		✓		
Cresci-17 [28]	http://mib.projects.iit.cnr.it/dataset.html	3,474 legitimate users, 10,894 fake bots, 6,637,616 tweets and relations.	✓		✓	✓	
Midterm-18 [174]	https://botometer.osome.iu.edu/bot-repository/datasets.html	8,092 legitimate users and 42,446 fake bots.	✓		✓	✓	
Twibot-22 [40]	https://github.com/LuoUndergradXJTU/TwiBot-22	860,057 legitimate users, 139,943 fake bots, 86,764,167 tweets and 170,185,937 relations.	✓		✓	✓	
MGTAB [140]	https://github.com/GraphDetec/MGTAB	7,554 legitimate users, 2,830 fake bots and 7 types of relations.	✓		✓	✓	
Ciao [144]	https://www.cse.msu.edu/~tangjili/trust.html	7,317 users, 283,319 ratings and 111,781 trust relations between users.	✓	✓			✓
Epinions [144]	https://www.cse.msu.edu/~tangjili/trust.html	22,164 users, 922,267 ratings and 355,754 trust relations between users.	✓	✓			✓

7.1.1 Connections, Relationships.

Trust can be inferred from user connections in the social network graph. Users in the same community tend to exhibit similar behaviors and levels of trust, and users who connect with trusted users are also likely to be considered more trustworthy. Trust features related to user connections or relationships include:

- **Node Connectivity:** number, strength, and type of connections (e.g., friends, followers, or mutual connections)
- **Centrality Measures:** degree centrality, betweenness centrality, eigenvector centrality, and closeness centrality
- **Community Detection:** inferring trust levels by identifying clusters or communities in social network graphs

7.1.2 Behaviors, Interactions, Transactions.

Trust relationships between users can be evaluated by analyzing their past behaviors, interactions, or

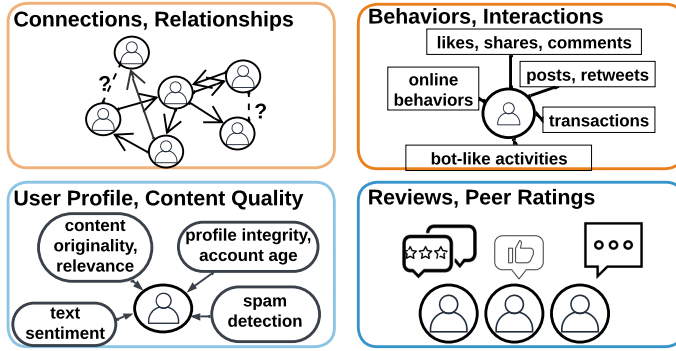


Fig. 5. Four commonly used datasets for calculating online trust.

transactions. Consistency in behavior can indicate trustworthiness, while inconsistency or sudden changes in behavior may indicate untrustworthiness. The frequency and type of interactions between users can also indicate trust levels. Generally speaking, trust can be measured by the following features:

- **Interaction Patterns:** frequency, consistency, and types of interactions (e.g., likes, shares, comments, messages, or group chats)
- **Behavior Patterns:** frequency, regularity, and timing of online behavior or social activities
- **Bot or Anomaly Detection:** the presence of automated, bot-like activities or abnormal behavior patterns

7.1.3 *User Characteristics, Content Quality.*

Analyzing user characteristics and user-generated content can also provide insights into user trust levels. Verified profiles and long-registered accounts are generally considered more trustworthy. High-quality, informative user-generated content is more likely to gain trust [6]. Representative trust features include:

- **User Profile:** profile integrity, account age, and consistency of personal information (i.e. IP address, name, bio, and profile picture)
- **Content Quality:** spelling, grammar, originality, coherence, informativeness, and relevance of content
- **Text Sentiment:** the sentiment of text content (e.g., positive, negative, or neutral), the emotional tone of user-generated content
- **Spam Detection:** the presence of spam, duplication, misinformation, disinformation, or manipulated content

7.1.4 *Reviews, Ratings.*

Trust models can also leverage ratings and reviews provided by users to evaluate user trustworthiness. There are two types of rating data. The first comes from online review platforms, where users provide feedback on the quality of services and products. The second type involves trust ratings, which users assign to their peers within the community. Trust features related to user-provided ratings or reviews include:

- **Product Review or Rating Behaviors:** rating consistency, rating distribution, review length, review frequency
- **Peer Endorsements:** reputation scores accumulated through interactions, aggregated peer ratings, endorsements, or badges

7.2 Promising Techniques for Building Trust Models

Various techniques and algorithms have been explored to build trust models. This section focuses on the research question: How is trust quantified? What algorithms can be used to build trust quantification models? In Table 3, we outline several representative trust modeling techniques, highlighting the unique perspective each technique brings to the field of trust modeling.

Table 3. Promising Techniques for Building Trust Models in Online Social Networks.

Techniques	Trust Modeling Mechanisms and Challenges	Instances
Reputation-Based	• User trustworthiness is measured through a reputation score, which is accumulated through behavior, contributions, and peer feedback during interactions. It is computationally transparent and widely applicable. • The challenge is that reputation trust models are vulnerable to malicious manipulation and Sybil attacks.	Rezvani and Rezvani [128] (2020), GOLI [65] (2023)
Probabilistic/ Bayesian	• Represent trust as a probability distribution and update trust using Bayesian inference or Beta distribution when new evidence emerges. They work well with sparse interactions or noisy data. • The problem is that modeling requires prior probabilities, and it is difficult to set prior probabilities without historical data.	Lingam et al. [94] (2020), Naumzik and Feuerriegel [117] (2022)
Hidden Markov Models (HMMs)	• Represent user trust as a hidden state (e.g., "low trust" or "building trust"). Transitions between trust states are computed based on the observed interactions using transition probabilities. It is suited to model sequences of hidden states and capture the temporal dynamics of trust. • The challenge is that HMMs require sufficient historical data to estimate the parameters.	ElSalamouny et al. [35] (2009), Liu and Datta [101] (2012)
Subjective Logic	• Represent trust relation as a combination of beliefs and uncertainties [belief b, disbelief d, uncertainty u], ($b, d, u \in [0, 1]$). Dempster-Shafer theory [80, 137, 172] can be used to combine the three parameters and compute a final trust value. It models the subjectivity and uncertainty inherent in trust relationships. • The definitions of opinion triplets and membership functions can vary in different contexts and may contain biases.	Cho et al. [24] (2019), Liu et al. [98] (2021)
Fuzzy Logic	• Represent trust as a linguistic variable with fuzzy sets (e.g., "high trust" or "medium trust"). It deals with ambiguity in trust and is valuable when trust cannot be modeled numerically. Additionally, multiple trust factors can be aggregated using operators such as "AND", "OR", and "NOT". • The difficulty lies in designing appropriate membership functions [58, 121] and fuzzy rules.	FTWDF-EEFA [104] (2021), Ma et al. [105] (2021)
Sentiment Analysis	• Inferring user trustworthiness by evaluating the emotional tone of user-generated content [95, 122]. Positive sentiment indicates trust, while negative sentiment indicates distrust. • However, these models often simplify emotions into binary or ternary categories, which may overlook nuanced emotions related to trust. Moreover, sentiment analysis often struggles to detect sarcasm or irony.	SQUAD [85] (2024), DDTrust [45] (2020), TextBlob [103] (2018)
Trust Game	• User interactions are modeled as repeated trust games, where individuals decide whether to trust or deceive based on potential rewards and risks [70, 143]. It is well-suited for market-oriented social networks where incentives matter. • Problems arise when users use pseudonyms or frequently change accounts, making it difficult to enforce long-term strategies or penalties for breaches of trust.	Hu et al. [59] (2021), Guo et al. [53] (2022)

continued on next page

Techniques	Trust Modeling Mechanisms and Challenges	Instances
Social Network Analysis (SNA)	• Represent the social network as a weighted graph, users as nodes, and trust relationships as weighted edges. Infer trust between users based on the network structure and connectivity. Common methods include graph theory and centrality measures (e.g., degree centrality, betweenness centrality, eigenvector centrality, and closeness centrality). * One challenge is that the model is vulnerable to malicious link attacks, and another challenge is the high computational cost in large-scale social networks.	Meo et al. [114] (2017), T-VLC [20] (2023), INTRUST [170] (2025)
Trust Propagation	• Infer trust relationships from connections and paths. Graph traversal algorithms such as breadth-first search and depth-first search can be used to build a complete trust network based on existing relationships. * Trust dilution along long paths. Moreover, trust evaluations may conflict when they are propagated from multiple sources, thereby requiring an appropriate aggregation mechanism.	AHNTF [169] (2024), Wu et al. [165] (2021), 3VSL-AT [98] (2021)
Graph Neural Networks (GNNs)	• Users are nodes, and each node aggregates trust-related signals from its neighbors to update its representation. After several layers of message passing, the trust score is predicted based on the final representation of the node. It can capture complex relational dependencies and is widely used in trust prediction and trust propagation. * Nevertheless, the complex structure of GNNs leads to limited model interpretability, and their performance is further limited by data sparsity.	TrustGNN [63] (2023), Gatrust [68] (2022), Trustguard [157] (2024)
Supervised Learning	• It leverages historical data to predict user trustworthiness in new interactions. First, it extracts relevant features from historical data, such as user behavior, interaction patterns, or text sentiment. Next, it uses supervised algorithms (e.g., SVMs, random forests, or neural networks) to learn the mapping between these features and labeled trust levels. * Yet, it requires a large amount of labeled data (trusted/untrusted) for training.	NeuralWalk [96] (2019), Guardian [93] (2020)
Community Detection	• Group users into communities based on similar trust behaviors or interaction patterns. It can detect hidden trust relationships between users, especially in the absence of explicit trust labels. It also helps identify and isolate malicious users or groups. * Applying community detection to large-scale social networks can be computationally expensive.	UnDBot [123] (2024), SBCE [94] (2020)
Anomaly Detection	• Identify unusual patterns in user behavior that deviate from normal. It establishes a baseline of anticipated activity to flag potential fraudulent behavior or security threats. Common techniques include statistical methods (e.g., Z-scores, 3-sigma rule) or machine learning methods (e.g., isolation forests, graph neural networks). * User behavior and interactions are evolving (concept drift). Therefore, adaptive or online learning models are needed to continuously update the trust or anomaly thresholds.	TA-Detector [162] (2024), THGNN [92] (2023)
Blockchain	• Blockchain enables decentralized trust computation, immutability of interactions, and transparency in trust relationships. Cryptographic protocols and consensus mechanisms can be used for trust evaluation. Smart contracts can automatically calculate and update trust scores. Tokenization and incentive mechanisms help promote honest and trustworthy behavior. * This approach is still in its early stages but holds potential.	Social-Chain [173] (2021), Yu et al. [175] (2023)
Cognitive Models	• Incorporate principles from psychology and cognitive science into trust models, as well as human factors such as emotional bonding, cognitive assessment, and behavioral analysis.	Karami et al. [76] (2021), CHASSIS [90] (2020)

Modern trust modeling techniques are built upon basic trust models. Next, we will explore several basic trust models and discuss the mechanisms they use to quantify trust in online social networks. We also provide the latest variants of these base models for further reference.

7.2.1 Reputation Trust Model.

Trust is inferred from the reputation score accumulated through multiple interactions. The reputation score of user i is calculated by aggregating the feedback provided by its peers, $R_i = \frac{\sum_{j \in S_i} t_{ij} \times \omega_j}{\sum_{j \in S_i} \omega_j}$, where S_i is the set of users that provide peer feedback about user i , t_{ij} is the feedback from user j to user i , and ω_j is the weight of the feedback from user j .

For example, the Beta reputation model [74] utilizes the Beta probability distribution to calculate trustworthiness. The trust evaluation value of A towards B is calculated as $T_{A,B} = \frac{\alpha+1}{\alpha+\beta+2}$, where α is the number of positive interactions and β is the number of negative interactions. Furthermore, if multiple users $C_1, C_2, \dots, C_n$ have interacted with B and provided trust evaluations, A can aggregate these trust evaluations to form an overall trust evaluation of B , $T_{A,B} = \sum_{i=1}^n T_{C_i,B} \times \omega_i$, where ω_i is the weight of the trust evaluation provided by C_i .

Reputation and trust models are commonly used in e-commerce and community-based applications, where peer feedback is crucial. State-of-the-art variants of reputation trust models include Rezvani and Rezvani [128] (2020) and GOLI [65] (2023).

7.2.2 EigenTrust Model.

Online social networks can be represented as weighted graphs, where users are nodes and trust relationships are weighted edges. EigenTrust [75] calculates a global trust score for each node by aggregating local trust values rated by peer nodes (Algorithm 1).

Algorithm 1: EigenTrust [75]

- 1 **Local Trust Values:** Nodes rate their peer nodes based on past interactions.
 - 2 **Normalization and Trust Matrix:** Local trust values are normalized to ensure they sum to 1 for each node. c_{ij} is the local trust value that node i assigns to node j and normalized to $s_{ij} = \frac{c_{ij}}{\sum_k c_{ik}}$.
The normalized local trust values form the trust matrix S , where $S_{ij} = s_{ij}$.
 - 3 **Global Trust Computation:** Compute the principal eigenvector of the trust matrix S using the power iteration method and obtain the global trust value vector $\mathbf{t}$.
 - 4 **Iterative Computation:** The global trust vector $\mathbf{t}$ is computed iteratively using Equation 1.
-

The global trust vector $\mathbf{t}$ is the solution to the equation, $\mathbf{t} = S^T \mathbf{t}$, subject to the normalization condition $\sum_i t_i = 1$ and $t_i \geq 0$ for all i , where t_i is the global trust value of node i . The algorithm iteratively updates the trust vector $\mathbf{t}$ until it converges,

$$\mathbf{t}^{(k+1)} = d S^T \mathbf{t}^{(k)} + (1 - d) \mathbf{t}^{(0)} \quad (1)$$

where d is the damping factor (typically around 0.85) [18] and $\mathbf{t}^{(0)}$ is the initial trust vector (usually set to a uniform distribution $\mathbf{e}$).

One of the latest variants of the EigenTrust model is Vedula et al. [152] (2017).

7.2.3 Trust Propagation Model.

According to the transitivity of trust [51], if A trusts B , and B trusts C , then A may also trust C to some extent.

$$T_{A,C} = T_{A,B} \times T_{B,C} \quad (2)$$

where $T_{A,B}$ represents the trust value of A towards B , $T_{B,C}$ represents the trust value of B towards C , and $T_{A,C}$ represents the trust value of A towards C , which is derived from $T_{A,B}$ and $T_{B,C}$.

State-of-the-art variants of trust propagation model include Wu et al. [165] (2021), AtNE-Trust [159] (2020), and 3VSL-AT [98] (2021).

7.2.4 TrustRank Model.

TrustRank is a link-analysis algorithm based on the PageRank algorithm [120], designed to combat web spam and low-quality content [55]. The core idea is that high-quality web pages will link to other high-quality web pages, while spam pages rarely receive links from trusted pages.

First, construct a web graph where web pages are nodes and hyperlinks are directed edges. TrustRank starts with a small set of manually verified trusted pages (trusted seeds) and propagates trust through outgoing links (Algorithm 2).

Algorithm 2: TrustRank [55]

- 1 **Selection of Seed Set:** Manually select a small set of highly trusted pages (nodes) as trusted seeds S .
 - 2 **Trust Vector Initialization:** T_0 is the initial trust vector. For each node in the seed set $n \in S$, $T_0(n) = 1$, otherwise $T_0(n) = 0$.
 - 3 **Iterative Computation:** Iteratively update the trust vector T until trust scores stabilize (Equation 3).
-

The transition probability matrix is P , where P_{mn} represents the probability of transitioning from node m to node n . If there is a link from m to n , then $P_{mn} = \frac{1}{Out(m)}$, where $Out(m)$ is the number of outgoing links from m . The trust vector T is iteratively calculated using the equation,

$$T^{(k+1)} = (1 - d) T_0 + d P^T T^{(k)} \quad (3)$$

where d is the damping factor (typically around 0.85) [18]. The algorithm converges when the change in trust values between two consecutive iterations falls below a small threshold, $\|T^{(k+1)} - T^{(k)}\| < \epsilon$. Generally, pages that are closer to trusted seeds (calculated by link distance) have higher trust scores.

State-of-the-art variants of PageRank-based trust models include MPR [184] and AHNTTP [169].

7.2.5 Subjective Logic Trust Model.

Subjective logic is a probabilistic logic framework that explicitly models uncertainty when expressing trust [22, 72, 73]. Trust is represented by an opinion tuple $\omega = (b, d, u, a)$, and $b + d + u = 1$. Belief (b) indicates the degree to which a proposition is believed to be true, disbelief (d) indicates the degree to which it is believed to be false, uncertainty (u) reflects the degree of uncertainty about its truth, and base rate (a) is the prior probability that the proposition is true without any evidence. Subjective logic has several operators for trust [73]:

- **Discounting Operator:** When A trust B with opinion $T_{A,B} = (b_{AB}, d_{AB}, u_{AB}, a_{AB})$, and B provides an opinion about the proposition x , $\omega_{B \rightarrow x} = (b_x, d_x, u_x, a_x)$. How much does A trust the proposition x ?

$$\omega_{A \rightarrow x} = T_{A,B} \otimes \omega_{B \rightarrow x} = (b_{AB} \cdot b_x, b_{AB} \cdot d_x, d_{AB} + u_{AB} + b_{AB} \cdot u_x, a_x) \quad (4)$$

where $\omega_{A \rightarrow x}$ is the opinion of A about the proposition x after discounting the opinion of B about the proposition x . Discounting is applying A's trust in B to B's opinion.

- **Consensus Operator (Fusing Opinions):** Aggregate multiple independent opinions. Assume that A and C both have an opinion about whether B is trustworthy ($\omega_{A,B}$ and $\omega_{C,B}$). Then the aggregated opinion is,

$$\omega_{A,B} \oplus \omega_{C,B} = \left(\frac{b_{A,B} \cdot u_{C,B} + b_{C,B} \cdot u_{A,B}}{u_{A,B} + u_{C,B} - u_{A,B} \cdot u_{C,B}}, \frac{d_{A,B} \cdot u_{C,B} + d_{C,B} \cdot u_{A,B}}{u_{A,B} + u_{C,B} - u_{A,B} \cdot u_{C,B}}, \frac{u_{A,B} \cdot u_{C,B}}{u_{A,B} + u_{C,B} - u_{A,B} \cdot u_{C,B}}, a_{A,B} \right) \quad (5)$$

where the base rate is assumed to be the same for both opinions, $a_{A,B} = a_{C,B}$.

State-of-the-art variants of subjective logic trust models include Cho et al. [24] (2019) and Liu et al. [98] (2021).

7.2.6 Fuzzy Trust Model.

Fuzzy logic deals with the ambiguity in trust using linguistic terms (e.g., "high trust" or "low trust") and membership functions [177, 178]. There are several key components:

- (1) **Fuzzy Sets and Membership Functions:** Suppose that a seller's trustworthiness is evaluated based on two factors: the number of positive reviews (Experience) and the average rating (Reputation). The input and output fuzzy sets can be defined as:
 - Input - Experience: {Low, Medium, High}, Reputation: {Poor, Average, Good}
 - Output - Trust: {Low, Medium, High}

Membership functions convert crisp inputs into fuzzy values [58, 121]. Here is an example of

a triangular membership function for "Low Experience": $\mu_{E_L}(x) = \begin{cases} 1 & \text{if } x \leq a \\ \frac{b-x}{a} & \text{if } a < x \leq b \\ 0 & \text{if } x > b \end{cases}$, where

x is the number of positive reviews, a and b are conditional thresholds, and $\mu_{E_L}(x) \in [0, 1]$. The membership functions for "Medium Experience" ($\mu_{E_M}(x)$) and "High Experience" ($\mu_{E_H}(x)$) are similar, except that the threshold conditions are different.

- (2) **Fuzzy Rules:** Define a set of fuzzy rules to map input fuzzy variables to output fuzzy variables.
 - Rule 1:** IF Experience is High AND Reputation is Good, THEN Trust is High.
 - Rule 2:** IF Experience is Medium AND Reputation is Average, THEN Trust is Medium.
 - Rule 3:** IF Experience is Low OR Reputation is Poor, THEN Trust is Low.
- (3) **Fuzzy Inference:** Apply fuzzy rules to calculate fuzzy output variables. For example, applying **Rule 1** using the Mamdani inference method: IF $\mu_{E_H}(x)$ AND $\mu_{R_G}(y)$ THEN $\mu_{T_H}(z) = \min(\mu_{E_H}(x), \mu_{R_G}(y))$.
- (4) **Defuzzification:** Combine the fuzzy outputs of all rules to form a single fuzzy set and convert the aggregated fuzzy set $\mu_T(z)$ into a single crisp trust value t_0 . For example, using the centroid method, $t_0 = \frac{\int z \cdot \mu_T(z) dz}{\int \mu_T(z) dz}$.

State-of-the-art variants of fuzzy trust models include FTWDF-EEFA [104] (2021) and Ma et al. [105] (2021).

7.2.7 Bayesian Trust Model.

Bayesian trust models use Bayesian inference [14] to update the probability of an user being trustworthy based on new evidence or interactions [25, 161]:

$$P(T_{A,B}^{\text{posterior}} | E) = \frac{P(E|T_{A,B}) \times P(T_{A,B}^{\text{prior}})}{P(E)} \quad (6)$$

where $T_{A,B}$ represents the degree of trust that A has in B, $P(T_{A,B}^{\text{prior}})$ is the prior probability, the initial trust value before any interaction or evidence, $P(E)$ is the probability of observing evidence E , $P(E|T_{A,B})$ is the probability of observing evidence E given the current trust value $T_{A,B}$, and $P(T_{A,B}^{\text{posterior}} | E)$ is the posterior probability of trust given evidence E .

State-of-the-art variants of Bayesian trust models include Cho et al. [24] (2019), Lingam et al. [94], Naumzik and Feuerriegel [117] (2022), 3VSL-AT [98] (2021), and T-VLC [20] (2023).

7.2.8 Game Theoretic Trust Model.

Game theory models trust as strategic interactions between rational players [70, 143]. Players evaluate whether to trust or deceive each other based on the potential rewards and risks in the trust game.

A well-known model is the repeated Prisoner's Dilemma [82, 125, 126]. The payoff matrix of a single-round Prisoner's Dilemma is:

Player A \ Player B	Trust	Defect
Trust	(R, R)	(S, T)
Defect	(T, S)	(P, P)

where R is the reward for mutual trust, S is the payoff of the player who chose to trust and was deceived, T is the temptation of deception, and P is the punishment for mutual deception. Typically, $T > R > P > S$.

Suppose player A trusts player B with probability T_{AB} , and the expected payoff of player A is E_A . If A chooses to cooperate, then $E_A(C) = T_{AB} \cdot R + (1 - T_{AB}) \cdot S$, otherwise, if A chooses to defect, then $E_A(D) = T_{AB} \cdot T + (1 - T_{AB}) \cdot P$. In the repeated trust game, A can adjust trust T_{AB} according to the difference between actual payoff and expected payoff:

$$T_{AB}^{(t+1)} = T_{AB}^{(t)} + \alpha(R - E_A) \quad (7)$$

where α is a learning rate.

State-of-the-art variants of game-theoretic trust models include PRADA-TF [106] (2024), Hu et al. [59] (2021), and Guo et al. [53] (2022).

7.3 What Application Scenarios Benefit from Trust Models

This section revolves around the question: What application scenarios benefit from trust quantification models? Below, we outline eight application scenarios of online social networks where trust models play an important role.

7.3.1 Building Connections and Community Formation.

An important role of trust models in online social networks is to measure, infer, and manage trust relationships between users.

- (1) **Building Connections:** Suggesting trustworthy users to follow or connect with based on past interactions, shared connections, and user reliability [36].
- (2) **Community Formation:** Facilitating the formation of trust-based communities or discussion groups where users are more likely to engage productively.

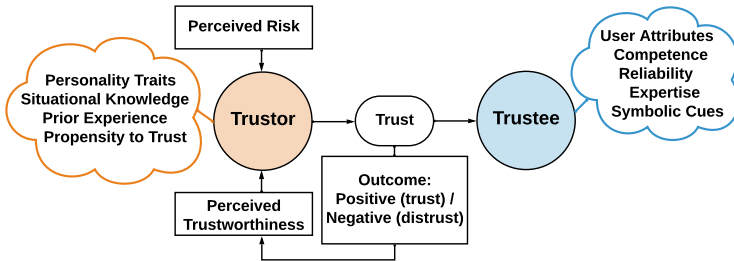


Fig. 6. Trust Formation Model [9, 10, 110]

Figure 6 illustrates the formation of trust between a trustor and a trustee [9, 10]. From the trustee's perspective, influencing factors include user attributes (e.g., age, gender, education, residence, income, and ethical standards), competence, reliability, expertise, and symbolic cues that indicate

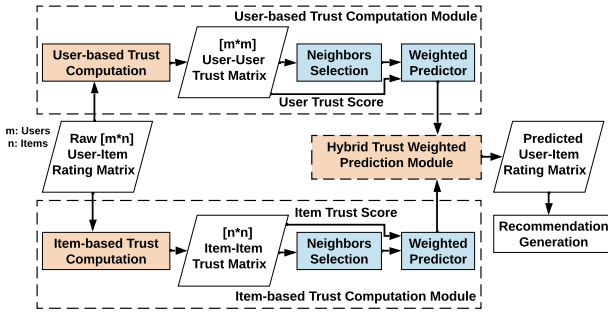


Fig. 7. Trust-Based Collaborative Filtering [138]

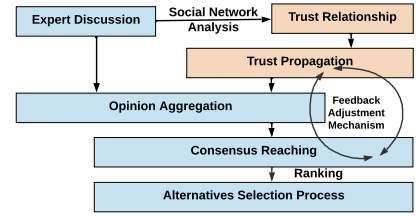


Fig. 8. Trust-Driven Group Decision-Making [151, 181]

trust. From the trustor's perspective, influencing factors include personality traits, situational knowledge, prior experience, and perceived trustworthiness.

7.3.2 Information Credibility and Fake News Detection.

Trust models assign trust scores to users based on their past behavior, expertise, and reputation. The credibility of information can be inferred from the trustworthiness of its source [2].

- (1) **Information Credibility Assessment:** Combining content features (e.g., linguistic style, sentiment, and fact-checking consistency) with trust features (e.g., user reliability and trusted endorsements) to measure information credibility. It helps filter out misleading information that appears linguistically credible but originates from low-trust accounts.
- (2) **Fake News Detection:** Combating fake news, rumors, harmful propaganda, and bot-driven campaigns [182].

7.3.3 Content Filtering and Recommendation.

Trust models distinguish and promote posts, articles, or media shared by trusted users when personalizing information feeds. Furthermore, trust models improve the accuracy of recommendation methods compared to those based solely on popularity or similarity.

- (1) **Content Filtering:** Filtering and ranking posts, articles, or videos based on the trustworthiness of the source.
- (2) **Product Recommendations:** Recommending products or services based on the preferences of trusted peers [188].

Trust-based collaborative filtering improves recommendation performance by prioritizing the opinions of trusted neighbors. As shown in Figure 7, the calculated trust scores are first applied to rank and select trusted neighbors and then used as weighting factors when aggregating their preferences. Trust models also help alleviate the data sparsity and cold start problems inherent in collaborative filtering [60].

7.3.4 Group Decision-Making and Consensus Reaching.

In scenarios where users collectively make decisions, trust models help prioritize and promote opinions from reliable sources. This filtered view enables more efficient decision-making and allows the group to focus on information from highly trusted sources.

As illustrated in Figure 8, group decision-making includes four main steps: expert discussion, opinion aggregation, consensus reaching, and alternative selection. Trust, as a motivating and weighting factor, is incorporated into the feedback and adjustment process and accelerates consensus reaching [100, 151, 181].

7.3.5 Online Collaboration and Crowdsourcing.

Trust models help identify reliable contributors based on past contributions, collaboration history, and peer reviews. Through trust scores, collaboration and crowdsourcing platforms can prioritize tasks to reliable contributors, thereby facilitating online collaboration and knowledge sharing.

7.3.6 Social Commerce and Transactions.

In social platforms with market functions, trust models help evaluate the reliability of sellers and enhance transaction confidence [150].

- (1) **Seller and Buyer Trust:** Building trust between buyers and sellers to enable peer-to-peer transactions [107].
- (2) **Secure Transactions:** Identifying fake or biased reviews and preventing fraud in transactions.

7.3.7 Influence Analysis and Information Diffusion.

Trust models can enhance the identification of influencers beyond just the number of followers. Selecting influencers based on trustworthiness can help correct or counteract misleading content, reduce the spread of rumors, and avoid the pitfalls of misinformation [91].

- (1) **Influencer Identification:** Ranking individuals by trust to identify authoritative voices or opinion leaders in communities [1, 65, 108].
- (2) **Information Diffusion:** Trust-aware information diffusion models help identify whether trending content is likely to be misinformation.

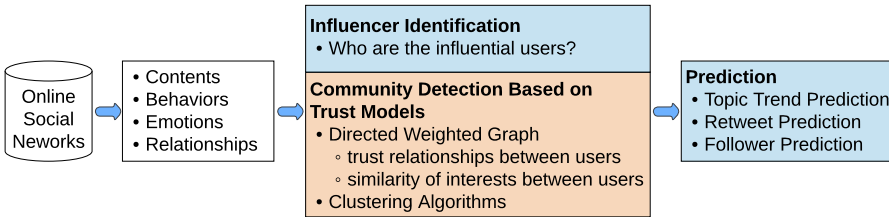


Fig. 9. Trust-Based Community Detection and Information Diffusion [91, 149]

Figure 9 shows a trust-based community detection and information diffusion model [91, 149]. Trust models were used for influencer identification and community detection, enhancing the accuracy of information diffusion prediction.

7.3.8 Security and Privacy.

Trust models help prevent fraud and support privacy protection, strengthening the security of online interactions [129].

- (1) **Fraud Detection:** Identifying malicious, fake, or sybil accounts and detecting potential fraud, phishing, or spam by evaluating unusual behavior patterns [94, 118, 123].
- (2) **Access Control:** Granting or restricting access to personal or sensitive information based on trust levels.

8 DISCUSSION AND FINDINGS

To answer the question, “What is trust? How is trust defined in online social networks,” Trust is a willingness to accept vulnerability based on positive expectations of others. It stems from both rational evaluation and emotional bonds. Trust takes time to build, especially in anonymous

online communities, but it can also be destroyed quickly. In online social network, trust can be defined as the confidence that users have in the reliability, integrity, and security of social networking platforms and their fellow users:

- **Reliability:** Confidence in the predictable behavior of fellow users and the consistency and reliability of online interactions.
- **Integrity:** Confidence in the authenticity of shared content, explainability of decision-making algorithms, and transparency in data use.
- **Competence:** Confidence in the capabilities of online platforms and the security against unauthorized access or cyber threats.

To answer the question, “What data is available to compute online trustworthiness?”, computing online trustworthiness involves analyzing various datasets and carefully selecting trust features to formulate effective trust models. Considering availability and practicality, we outline four dataset categories and their corresponding trust-related features. The most commonly used trust features are: (1) the number and strength of connections, (2) the frequency and consistency of interactions, (3) the quality and relevance of shared content, and (4) reputation scores established through peer review or endorsement. In short, ensuring data quality and selecting the most relevant trust features are key to effective trust modeling in online social networks.

To answer the question, “How is online trust quantified?”, we categorize the latest trust quantification models into ten algorithmic categories. In reputation-based trust models, trust is quantified as a numerical score accumulated and aggregated from past behaviors, including interactions, ratings, or endorsements from peers. Probabilistic trust models represent trust as a probability distribution and utilize Bayesian inference to update posterior trust beliefs with new evidence [83, 161]. Subjective logic and fuzzy logic deal with uncertainty and ambiguity in trust relationships [72, 121]. Fuzzy logic models express trust levels using linguistic terms such as “low”, “medium”, and “high” and update them according to fuzzy rules [177, 178]. Moreover, graph-based trust models treat the social network as a graph where nodes are users and edges carry trust scores [69]. Trust is inferred from connections and can be aggregated along multiple paths (e.g., weighted average of neighbors’ trust). Machine learning models exploit features of interactions and network structure to learn trustworthiness and can model nonlinear patterns [156]. For example, graph neural networks learn trust embeddings directly from social graphs. In addition, game theoretic models model trust as the outcome of strategic interactions. Hybrid models integrate social graph features, peer ratings, and behavioral data into one trust score. Although these trust models originate from different modeling perspectives, they collectively deepen our understanding of trust and transform trust from an abstract concept into a concrete numerical variable in online social networks.

To answer the question, “What application scenarios benefit from trust quantification models?”, three have been identified. First, trust models help form a vibrant and supportive community. Trust models help foster trust-based communities or discussion groups, where users are more likely to engage in productive and effective interactions. Additionally, trust models can identify reliable and influential users who can act as moderators to guide community interactions and encourage the dissemination of accurate, high-quality content. Second, trust models help maintain content quality and support group decision-making and crowdsourcing. Trust can be used as a screening criterion to filter out misleading or harmful content, while prioritizing reliable content [141, 187]. Moreover, trust scores can be used as weights in opinion aggregation or crowdsourcing, facilitating collaboration or group decision-making. Third, trust models can enhance the security of online social networks. They help detect fake profiles and malicious users, thereby reducing the risk

of fraud or deception. By analyzing user behavior patterns and social connections, trust models can identify and prevent suspicious activities or fraudulent behavior early on.

9 OPEN CHALLENGES AND RESEARCH DIRECTIONS

Trust modeling in online social networks is challenging due to data sparsity, model bias, evolving trust, privacy issues, and malicious manipulation. In this section, we highlight unresolved challenges and suggest potential directions for progress.

9.1 Data Sparsity and Model Bias

Trust often has to be inferred from partial or noisy data, as trust information is rarely complete due to missing profiles, private interactions, etc. Also, trust modeling suffer from data sparsity and cold start problems. It is difficult to calculate a reliable trust score for new users or users with limited interactions. Moreover, bias and unfairness in trust models [134] may unfairly favor or penalize certain demographic groups.

9.2 Complex and Evolving Trust Relationships

First, trust is context-dependent and varies by topic, time, or situation (e.g., someone may be trusted for movie recommendations but not for financial advice). Additionally, trust changes based on user behavior, interactions, and new information. Trust models must adapt to these changes without being outdated or biased by past interactions. Second, users interact across multiple online platforms. Therefore, building a cross-platform trust model and deriving an overall trust score is necessary, although difficult.

9.3 Privacy Concerns and Malicious Manipulation

Collecting and analyzing trust-related data may expose sensitive user information. Therefore, striking a balance between trust modeling and user privacy protection is crucial [129]. In addition, detecting fake accounts, bots [94, 123], and mitigating Sybil attacks [118] or reputation manipulation are ongoing challenges.

10 CONCLUSIONS

Trust modeling research in online social networks focuses on developing computational models and algorithms to simulate the formation, maintenance, or erosion of trust in online communities. Going further, it helps assess trust relationships, verify the credibility of information, detect malicious or spam activity, and foster more trustworthy and high-quality online communities. To define trust, trust is a willingness to accept vulnerability based on positive expectations of others. It stems from both rational evaluation and emotional bonds. Trust takes time to build, especially in anonymous online communities, but can be lost quickly. To simplify and facilitate trust modeling, we dissect the concept of trust and identify several factors that influence the formation and evolution of online trust, including competence, reciprocity, consistency, transparency, and similarity. Next, we systematically classify state-of-the-art trust quantification models into ten algorithmic categories. Various trust models collectively advance our understanding of how trust is established, maintained, or propagated. Subjective logic and fuzzy logic deal with the inherent uncertainty and ambiguity in trust relationships. Reputation-based trust models aggregate community opinions and accumulate trust scores based on behavior, contributions, and peer feedback during interactions. Graph-based trust models simulate social networks as weighted graphs, where users are nodes and trust relationships are edges. Trust is propagated or aggregated through connections and paths. Context-aware trust models are developed for specific domains, topics, or situations, rather than using a one-size-fits-all approach.

Research on trust models in online social networks is still in its emerging stage, and many issues remain unresolved. As part of our efforts, we provide an implementation-centric handbook on trust modeling and suggest potential research directions to encourage further exploration in this field. While building effective trust models is challenging, we believe that through in-depth research, trust can evolve from an ideal concept into a powerful tool in online interactions.

REFERENCES

- [1] Mohammed Ali Al-Garadi, Kasturi Dewi Varathan, Sri Devi Ravana, Ejaz Ahmed, Ghulam Mujtaba, Muhammad Usman Shahid Khan, and Samee U Khan. 2018. Analysis of online social network connections for identification of influential users: Survey and open research issues. *ACM Computing Surveys (CSUR)* 51, 1 (2018), 1–37.
- [2] Majed Alrubaian, Muhammad Al-Qurishi, Mohammad Mehedi Hassan, and Atif Alamri. 2016. A credibility analysis system for assessing information on twitter. *IEEE Transactions on Dependable and Secure Computing* 15, 4 (2016), 661–674.
- [3] Solomon E Asch. 2016. Effects of group pressure upon the modification and distortion of judgments. In *Organizational influence processes*. Routledge, 295–303.
- [4] Mojgan Askarizadeh, Behrouz Tork Ladani, and Mohammad Hossein Manshaei. 2019. An evolutionary game model for analysis of rumor propagation and control in social networks. *Physica A: statistical mechanics and its applications* 523 (2019), 21–39.
- [5] Marius Becherer, Omar Khadeer Hussain, Yu Zhang, Frank den Hartog, and Elizabeth Chang. 2024. On Trust Recommendations in the Social Internet of Things—A Survey. *Comput. Surveys* 56, 6 (2024), 1–35.
- [6] Amin Beheshti, Vahid Moraveji-Hashemi, Shahpar Yakhchi, Hamid Reza Motahari-Nezhad, Seyed Mohssen Ghafari, and Jian Yang. 2020. personality2vec: Enabling the analysis of behavioral disorders in social networks. In *Proceedings of the 13th international conference on web search and data mining*. 825–828.
- [7] Juan Benet. 2014. Ipfs-content addressed, versioned, p2p file system. *arXiv preprint arXiv:1407.3561* (2014).
- [8] Peter M Blau. 1964. Justice in social exchange. *Sociological inquiry* 34, 2 (1964).
- [9] Bernd Blöbaum. 2016. Key factors in the process of trust. On the analysis of trust under digital conditions. *Trust and communication in a digitized world: Models and concepts of trust research* (2016), 3–25.
- [10] Bernd Blöbaum. 2021. Some thoughts on the nature of trust: Concept, models and theory. In *Trust and Communication: Findings and Implications of Trust Research*. Springer, 3–28.
- [11] Marián Boguná, Romualdo Pastor-Satorras, Albert Díaz-Guilera, and Alex Arenas. 2004. Models of social networks based on social distance attachment. *Physical review E* 70, 5 (2004), 056122.
- [12] Tilman Börgers. 2015. *An introduction to the theory of mechanism design*. Oxford University Press, USA.
- [13] George EP Box and Norman R Draper. 1987. *Empirical model-building and response surfaces*. John Wiley & Sons.
- [14] George EP Box and George C Tiao. 2011. *Bayesian inference in statistical analysis*. John Wiley & Sons.
- [15] Ryan L Boyd, Ashwini Ashokkumar, Sarah Seraj, and James W Pennebaker. 2022. The development and psychometric properties of LIWC-22. *Austin, TX: University of Texas at Austin* 10 (2022).
- [16] Diego De Siqueira Braga, Marco Niemann, Bernd Hellingrath, and Fernando Buarque De Lima Neto. 2018. Survey on computational trust and reputation models. *ACM Computing Surveys (CSUR)* 51, 5 (2018), 1–40.
- [17] Alain Bretto. 2013. Hypergraph theory. *An introduction. Mathematical Engineering. Cham: Springer* 1 (2013).
- [18] Sergey Brin and Lawrence Page. 1998. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems* 30, 1-7 (1998), 107–117.
- [19] Christiano Castelfranchi and Rino Falcone. 2010. *Trust theory: A socio-cognitive and computational model*. John Wiley & Sons.
- [20] Luis Chamba-Eras, Ana Arruarte, and Jon A Elorriaga. 2023. T-VLC: A Trust Model for Virtual Learning Communities. *IEEE Transactions on Learning Technologies* 16, 5 (2023), 847–860.
- [21] In-Koo Cho and David M Kreps. 1987. Signaling games and stable equilibria. *The Quarterly Journal of Economics* 102, 2 (1987), 179–221.
- [22] Jin-Hee Cho. 2018. Dynamics of uncertain and conflicting opinions in social networks. *IEEE transactions on computational social systems* 5, 2 (2018), 518–531.
- [23] Jin-Hee Cho, Kevin Chan, and Sibel Adali. 2015. A survey on trust modeling. *ACM Computing Surveys (CSUR)* 48, 2 (2015), 1–40.
- [24] Jin-Hee Cho, Scott Rager, John O'Donovan, Sibel Adali, and Benjamin D Horne. 2019. Uncertainty-based false information propagation in social networks. *ACM Transactions on Social Computing* 2, 2 (2019), 1–34.
- [25] Zi Chu, Steven Gianvecchio, Haining Wang, and Sushil Jajodia. 2012. Detecting automation of twitter accounts: Are you a human, bot, or cyborg? *IEEE Transactions on dependable and secure computing* 9, 6 (2012), 811–824.

- [26] Robert B Cialdini and Noah J Goldstein. 2004. Social influence: Compliance and conformity. *Annu. Rev. Psychol.* 55, 1 (2004), 591–621.
- [27] Andy Clark. 2015. Radical predictive processing. *The Southern Journal of Philosophy* 53 (2015), 3–27.
- [28] Stefano Cresci, Roberto Di Pietro, Marinella Petrocchi, Angelo Spognardi, and Maurizio Tesconi. 2017. The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race. In *Proceedings of the 26th international conference on world wide web companion*. 963–972.
- [29] Cristobald De Kerchove and Paul Van Dooren. 2010. Iterative filtering in reputation systems. *SIAM J. Matrix Anal. Appl.* 31, 4 (2010), 1812–1834.
- [30] Morton Deutsch. 1958. Trust and suspicion. *Journal of conflict resolution* 2, 4 (1958), 265–279.
- [31] Morton Deutsch and Harold B Gerard. 1955. A study of normative and informational social influences upon individual judgment. *The journal of abnormal and social psychology* 51, 3 (1955), 629.
- [32] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
- [33] Nicholas DiFonzo and Prashant Bordia. 2007. *Rumor psychology: Social and organizational approaches*. American Psychological Association.
- [34] Brian Dowtin, Kofi Kyei11, Yasmin Eady, AM Jackson, Bernard Aldrich, Joseph Shelton, and Albert Esterline. 2024. A Notion of Trustworthiness Based on Centrality in a Social Network. *Human Factors in Cybersecurity* (2024), 52.
- [35] Ehab ElSalamouny, Vladimiro Sassone, and Mogens Nielsen. 2009. HMM-based trust model. In *International Workshop on Formal Aspects in Security and Trust*. Springer, 21–35.
- [36] Christian Esposito, Vincenzo Moscato, and Giancarlo Sperli. 2021. Trustworthiness assessment of users in social reviewing systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 52, 1 (2021), 151–165.
- [37] Rino Falcone and Cristiano Castelfranchi. 2001. Social trust: A cognitive approach. *Trust and deception in virtual societies* (2001), 55–90.
- [38] Xinxin Fan, Ling Liu, Rui Zhang, Quanliang Jing, and Jingping Bi. 2020. Decentralized trust management: Risk analysis and trust aggregation. *ACM Computing Surveys (CSUR)* 53, 1 (2020), 1–33.
- [39] Lauren Fell, Andrew Gibson, Peter Bruza, and Pamela Hoyte. 2020. Human information interaction and the cognitive predicting theory of trust. In *Proceedings of the 2020 conference on human information interaction and retrieval*. 145–152.
- [40] Shangbin Feng, Zhaoxuan Tan, Herun Wan, Ningnan Wang, Zilong Chen, Binchi Zhang, Qinghua Zheng, Wenqian Zhang, Zhenyu Lei, Shujie Yang, et al. 2022. TwiBot-22: Towards graph-based Twitter bot detection. *Advances in Neural Information Processing Systems* 35 (2022), 35254–35269.
- [41] Lidia Fotia, Flávia Delicato, and Giancarlo Fortino. 2023. Trust in edge-based internet of things architectures: state of the art and research challenges. *Comput. Surveys* 55, 9 (2023), 1–34.
- [42] Xiaofeng Gao, Wenyi Xu, Mingding Liao, and Guihai Chen. 2021. Trust prediction for online social networks with integrated time-aware similarity. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 15, 6 (2021), 1–30.
- [43] Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin. 1995. *Bayesian data analysis*. Chapman and Hall/CRC.
- [44] Seyed Mohssen Ghafari, Amin Beheshti, Aditya Joshi, Cecile Paris, Adnan Mahmood, Shahpar Yakhchi, and Mehmet A Orgun. 2020. A survey on trust prediction in online social networks. *IEEE Access* 8 (2020), 144292–144309.
- [45] Seyed Mohssen Ghafari, Amin Beheshti, Aditya Joshi, Cecile Paris, Shahpar Yakhchi, Alireza Jolfaei, and Mehmet A Orgun. 2020. A dynamic deep trust prediction approach for online social networks. In *Proceedings of the 18th international conference on advances in mobile computing & multimedia*. 11–19.
- [46] Seyed Mohssen Ghafari, Aditya Joshi, Amin Beheshti, Cecile Paris, Shahpar Yakhchi, and Mehmet Orgun. 2019. DCAT: A deep context-aware trust prediction approach for online social networks. In *Proceedings of the 17th International Conference on Advances in Mobile Computing & Multimedia*. 20–27.
- [47] Anthony Giddens. 1990. *The consequence of modernity*. Stanford University Press.
- [48] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2020. Generative adversarial networks. *Commun. ACM* 63, 11 (2020), 139–144.
- [49] Sonja Grabner-Kräuter and Sofie Bitter. 2015. Trust in online social networks: A multifaceted perspective. In *Forum for social economics*, Vol. 44. Taylor & Francis, 48–68.
- [50] Jones Granatyr, Nardine Osman, João Dias, Maria Augusta Silveira Netto Nunes, Judith Masthoff, Fabricio Enembreck, Otto Robert Lessing, Carles Sierra, Ana Maria Paiva, and Edson Emilio Scalabrin. 2017. The need for affective trust applied to trust and reputation models. *ACM Computing Surveys (CSUR)* 50, 4 (2017), 1–36.
- [51] Ramanathan Guha, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. 2004. Propagation of trust and distrust. In *Proceedings of the 13th international conference on World Wide Web*. 403–412.
- [52] Zhen Guo, Jin-Hee Cho, and Chang-Tien Lu. 2022. Mitigating influence of disinformation propagation using uncertainty-based opinion interactions. *IEEE Transactions on Computational Social Systems* 10, 2 (2022), 435–447.

- [53] Zhen Guo, Jaber Valinejad, and Jin-Hee Cho. 2022. Effect of disinformation propagation on opinion dynamics: A game theoretic approach. *IEEE Transactions on Network Science and Engineering* 9, 5 (2022), 3775–3790.
- [54] Akshit Gupta, Debadeep Basu, Ramya Ghantasala, Sihang Qiu, and Ujwal Gadiraju. 2022. To trust or not to trust: How a conversational interface affects trust in a decision support system. In *Proceedings of the ACM Web Conference 2022*. 3531–3540.
- [55] Zoltán Gyöngyi, Hector Garcia-Molina, and Jan Pedersen. 2004. Combating web spam with trustrank. In *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*. 576–587.
- [56] Omar Hasan, Lionel Brunie, and Elisa Bertino. 2022. Privacy-preserving reputation systems based on blockchain and other cryptographic building blocks: A survey. *ACM Computing Surveys (CSUR)* 55, 2 (2022), 1–37.
- [57] Alan G Hawkes. 1971. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58, 1 (1971), 83–90.
- [58] Tzung-Pei Hong and Chai-Ying Lee. 1996. Induction of fuzzy rules and membership functions from training examples. *Fuzzy sets and Systems* 84, 1 (1996), 33–47.
- [59] Zhengyang Hu, Xiaopeng Li, Juan Wang, Chengyi Xia, Zhen Wang, and Matjaž Perc. 2021. Adaptive reputation promotes trust in social networks. *IEEE Transactions on Network Science and Engineering* 8, 4 (2021), 3087–3098.
- [60] Zhengdi Hu, Guangquan Xu, Xi Zheng, Jiang Liu, Zhangbing Li, Quan Z Sheng, Wenjuan Lian, and Hequn Xian. 2020. SSL-SVD: semi-supervised learning-based sparse trust recommendation. *ACM Transactions on Internet Technology (TOIT)* 20, 1 (2020), 1–20.
- [61] Teng-Chieh Huang, Razieh Nokhbeh Zaeem, and K Suzanne Barber. 2019. 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)* 19, 4 (2019), 1–20.
- [62] Teng-Chieh Huang, Razieh Nokhbeh Zaeem, and K Suzanne Barber. 2021. Identifying real-world credible experts in the financial domain. *Digital Threats: Research and Practice* 2, 2 (2021), 1–14.
- [63] Cuiying Huo, Dongxiao He, Chundong Liang, Di Jin, Tie Qiu, and Lingfei Wu. 2023. TrustGNN: Graph Neural Network-Based Trust Evaluation via Learnable Propagative and Composable Nature. *IEEE Transactions on Neural Networks and Learning Systems* (2023).
- [64] Brett W Israelsen and Nisar R Ahmed. 2019. “Dave... I can assure you... that it’s going to be all right...” A definition, case for, and survey of algorithmic assurances in human-autonomy trust relationships. *ACM Computing Surveys (CSUR)* 51, 6 (2019), 1–37.
- [65] Lokesh Jain, Rahul Katarya, and Shelly Sachdeva. 2023. Opinion leaders for information diffusion using graph neural network in online social networks. *ACM Transactions on the Web* 17, 2 (2023), 1–37.
- [66] Jinyuan Jia, Binghui Wang, and Neil Zhenqiang Gong. 2017. Random walk based fake account detection in online social networks. In *2017 47th annual IEEE/IFIP international conference on dependable systems and networks (DSN)*. IEEE, 273–284.
- [67] Le Jiang and Xinglin Zhang. 2019. BCOSN: A blockchain-based decentralized online social network. *IEEE Transactions on Computational Social Systems* 6, 6 (2019), 1454–1466.
- [68] Nan Jiang, Jie Wen, Jin Li, Ximeng Liu, and Di Jin. 2022. Gatrust: A multi-aspect graph attention network model for trust assessment in osns. *IEEE Transactions on Knowledge and Data Engineering* 35, 6 (2022), 5865–5878.
- [69] Wenjun Jiang, Guojun Wang, Md Zakirul Alam Bhuiyan, and Jie Wu. 2016. Understanding graph-based trust evaluation in online social networks: Methodologies and challenges. *Acm Computing Surveys (Csur)* 49, 1 (2016), 1–35.
- [70] Noel D Johnson and Alexandra A Mislin. 2011. Trust games: A meta-analysis. *Journal of economic psychology* 32, 5 (2011), 865–889.
- [71] Dominic Zaun Eu Jones and Eshwar Chandrasekharan. 2024. Measuring Epistemic Trust: Towards a New Lens for Democratic Legitimacy, Misinformation, and Echo Chambers. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW2 (2024), 1–33.
- [72] Audun Jøsang. 2016. *Subjective logic*. Vol. 3. Springer.
- [73] Audun Jøsang, Ross Hayward, and Simon Pope. 2006. Trust network analysis with subjective logic. In *Conference Proceedings of the Twenty-Ninth Australasian Computer Science Conference (ACSW 2006)*. Australian Computer Society, 85–94.
- [74] Audun Jøsang and Roslan Ismail. 2002. The beta reputation system. In *Proceedings of the 15th bled electronic commerce conference*, Vol. 5. Citeseer, 2502–2511.
- [75] Sepandar D Kamvar, Mario T Schlosser, and Hector Garcia-Molina. 2003. The eigentrust algorithm for reputation management in p2p networks. In *Proceedings of the 12th international conference on World Wide Web*. 640–651.
- [76] Mansoor Karami, Tahora H Nazer, and Huan Liu. 2021. Profiling fake news spreaders on social media through psychological and motivational factors. In *Proceedings of the 32nd ACM conference on hypertext and social media*. 225–230.

- [77] Davinder Kaur, Suleyman Uslu, Kaley J Rittichier, and Arjan Durrezi. 2022. Trustworthy artificial intelligence: a review. *ACM Computing Surveys (CSUR)* 55, 2 (2022), 1–38.
- [78] Diederik P Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).
- [79] Jon M Kleinberg. 1999. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)* 46, 5 (1999), 604–632.
- [80] Jürg Kohlas and Paul-André Monney. 2013. *A mathematical theory of hints: An approach to the Dempster-Shafer theory of evidence*. Vol. 425. Springer Science & Business Media.
- [81] Andrea Krausman, Catherine Neubauer, Daniel Forster, Shan Lakhmani, Anthony L Baker, Sean M Fitzhugh, Gregory Gremillion, Julia L Wright, Jason S Metcalfe, and Kristin E Schaefer. 2022. Trust measurement in human-autonomy teams: Development of a conceptual toolkit. *ACM Transactions on Human-Robot Interaction (THRI)* 11, 3 (2022), 1–58.
- [82] David M Kreps, Paul Milgrom, John Roberts, and Robert Wilson. 1982. Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic theory* 27, 2 (1982), 245–252.
- [83] Ugur Kuter and Jennifer Golbeck. 2010. Using probabilistic confidence models for trust inference in web-based social networks. *ACM Transactions on Internet Technology (TOIT)* 10, 2 (2010), 1–23.
- [84] Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *International conference on machine learning*. PMLR, 1188–1196.
- [85] Anastasios Lepipas, Anastasia Borovykh, and Soteris Demetriou. 2024. Username squatting on online social networks: A study on x. In *Proceedings of the 19th ACM Asia Conference on Computer and Communications Security*. 621–637.
- [86] Jure Leskovec and Julian McAuley. 2012. Learning to discover social circles in ego networks. *Advances in neural information processing systems* 25 (2012).
- [87] Roy J Lewicki and Barbara Benedict Bunker. 1995. *Trust in relationships: A model of development and decline*. Jossey-Bass/Wiley.
- [88] Bo Li, Peng Qi, Bo Liu, Shuai Di, Jingen Liu, Jiquan Pei, Jinfeng Yi, and Bowen Zhou. 2023. Trustworthy AI: From principles to practices. *Comput. Surveys* 55, 9 (2023), 1–46.
- [89] Dong Li, Zhiming Xu, Yishu Luo, Sheng Li, Anika Gupta, Katia Sycara, Shengmei Luo, Lei Hu, and Hong Chen. 2013. Modeling information diffusion over social networks for temporal dynamic prediction. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*. 1477–1480.
- [90] Hui Li, Hui Li, and Sourav S Bhowmick. 2020. Chassis: Conformity meets online information diffusion. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*. 1829–1840.
- [91] Mei Li, Xiang Wang, Kai Gao, and Shanshan Zhang. 2017. A survey on information diffusion in online social networks: Models and methods. *Information* 8, 4 (2017), 118.
- [92] Yilin Li, Jiaqi Zhu, Congcong Zhang, Yi Yang, Jiawen Zhang, Ying Qiao, and Hongan Wang. 2023. THGNN: An Embedding-based Model for Anomaly Detection in Dynamic Heterogeneous Social Networks. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. 1368–1378.
- [93] Wanyu Lin, Zhaolin Gao, and Baochun Li. 2020. Guardian: Evaluating trust in online social networks with graph convolutional networks. In *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*. IEEE, 914–923.
- [94] Greeshma Lingam, Rashmi Ranjan Rout, Durvasula VLN Somayajulu, and Sajal K Das. 2020. 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*. 708–718.
- [95] Bing Liu. 2022. *Sentiment analysis and opinion mining*. Springer Nature.
- [96] Guangchi Liu, Chenyu Li, and Qing Yang. 2019. Neuralwalk: Trust assessment in online social networks with neural networks. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 1999–2007.
- [97] Guangchi Liu, Qing Yang, Honggang Wang, Xiaodong Lin, and Mike P Wittie. 2014. Assessment of multi-hop interpersonal trust in social networks by three-valued subjective logic. In *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*. IEEE, 1698–1706.
- [98] Guangchi Liu, Qing Yang, Honggang Wang, and Alex X Liu. 2021. Trust Assessment in Online Social Networks. *IEEE Transactions on Dependable and Secure Computing* 18, 2 (2021), 994–1007.
- [99] Haochen Liu, Yiqi Wang, Wenqi Fan, Xiaorui Liu, Yaxin Li, Shaili Jain, Yunhao Liu, Anil Jain, and Jiliang Tang. 2022. Trustworthy ai: A computational perspective. *ACM Transactions on Intelligent Systems and Technology* 14, 1 (2022), 1–59.
- [100] Peide Liu, Yueyuan Li, and Peng Wang. 2022. Opinion dynamics and minimum adjustment-driven consensus model for multi-criteria large-scale group decision making under a novel social trust propagation mechanism. *IEEE Transactions on Fuzzy Systems* 31, 1 (2022), 307–321.
- [101] Xin Liu and Anwitaman Datta. 2012. Modeling context aware dynamic trust using hidden markov model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 26. 1938–1944.
- [102] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*

- (2019).
- [103] Steven Loria et al. 2018. textblob Documentation. *Release 0.15 2*, 8 (2018), 269.
 - [104] Zhihan Lv and Houbing Song. 2021. Trust mechanism of feedback trust weight in multimedia network. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 17, 4 (2021), 1–26.
 - [105] Xiujuan Ma, Zaiwu Gong, Guo Wei, and Enrique Herrera-Viedma. 2021. A new consensus model based on trust interactive weights for intuitionistic group decision making in social networks. *IEEE Transactions on Cybernetics* 52, 12 (2021), 13106–13119.
 - [106] Yash Mahajan, Jin-Hee Cho, and Ing-Ray Chen. [n. d.]. Privacy-Preserving and Diversity-Aware Trust-based Team Formation in Online Social Networks. *ACM Transactions on Intelligent Systems and Technology* ([n. d.]).
 - [107] Tatiana Marceda Bach, Wesley Vieira da Silva, Adriano Mendonça Souza, Claudineia Kudlawicz-Franco, and Claudimar Pereira da Veiga. 2020. Online customer behavior: perceptions regarding the types of risks incurred through online purchases. *Palgrave Communications* 6, 1 (2020), 1–12.
 - [108] Ericsson Marin, Ruocheng Guo, and Paulo Shakarian. 2020. Measuring time-constrained influence to predict adoption in online social networks. *ACM Transactions on Social Computing* 3, 3 (2020), 1–26.
 - [109] Paolo Massa, Martino Salvetti, and Danilo Tomasoni. 2009. Bowling alone and trust decline in social network sites. In *2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing*. IEEE, 658–663.
 - [110] Roger C Mayer, James H Davis, and F David Schoorman. 1995. An integrative model of organizational trust. *Academy of management review* 20, 3 (1995), 709–734.
 - [111] Daniel J McAllister. 1995. Affect and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of management journal* 38, 1 (1995), 24–59.
 - [112] D Harrison McKnight, Larry L Cummings, and Norman L Chervany. 1998. Initial trust formation in new organizational relationships. *Academy of Management review* 23, 3 (1998), 473–490.
 - [113] Pasquale De Meo. 2019. Trust prediction via matrix factorisation. *ACM Transactions on Internet Technology (TOIT)* 19, 4 (2019), 1–20.
 - [114] Pasquale De Meo, Katarzyna Musiał-Gabrys, Domenico Rosaci, Giuseppe ML Sarne, and Lora Aroyo. 2017. Using centrality measures to predict helpfulness-based reputation in trust networks. *ACM Transactions on Internet Technology (TOIT)* 17, 1 (2017), 1–20.
 - [115] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).
 - [116] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems* 26 (2013).
 - [117] Christof Naumzik and Stefan Feuerriegel. 2022. Detecting false rumors from retweet dynamics on social media. In *Proceedings of the ACM web conference 2022*. 2798–2809.
 - [118] James Newsome, Elaine Shi, Dawn Song, and Adrian Perrig. 2004. The sybil attack in sensor networks: analysis & defenses. In *Proceedings of the 3rd international symposium on Information processing in sensor networks*. 259–268.
 - [119] Noam Nisan and Amir Ronen. 1999. Algorithmic mechanism design. In *Proceedings of the thirty-first annual ACM symposium on Theory of computing*. 129–140.
 - [120] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. *The PageRank citation ranking: Bringing order to the web*. Technical Report. Stanford infolab.
 - [121] Yingnan Pan, Qi Li, Hongjing Liang, and Hak-Keung Lam. 2021. A novel mixed control approach for fuzzy systems via membership functions online learning policy. *IEEE Transactions on Fuzzy Systems* 30, 9 (2021), 3812–3822.
 - [122] Bo Pang, Lillian Lee, et al. 2008. Opinion mining and sentiment analysis. *Foundations and Trends® in information retrieval* 2, 1–2 (2008), 1–135.
 - [123] Hao Peng, Jingyun Zhang, Xiang Huang, Zhifeng Hao, Angsheng Li, Zhengtao Yu, and Philip S Yu. 2024. Unsupervised social bot detection via structural information theory. *arXiv preprint arXiv:2404.13595* (2024).
 - [124] James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of LIWC2015. (2015).
 - [125] William Poundstone. 2011. *Prisoner's dilemma*. Anchor.
 - [126] Arotol Rapoport. 1965. Prisoner's Dilemma: a Study in Conflict and Cooperation.
 - [127] Jakob Gulddahl Rasmussen. 2013. Bayesian inference for Hawkes processes. *Methodology and Computing in Applied Probability* 15 (2013), 623–642.
 - [128] Mohsen Rezvani and Mojtava Rezvani. 2020. A randomized reputation system in the presence of unfair ratings. *ACM Transactions on Management Information Systems (TMIS)* 11, 1 (2020), 1–16.
 - [129] Andrey Rodrigues, Maria Lúcia Villela, and Eduardo Feitosa. 2024. A Systematic Mapping Study on Social Network Privacy: Threats and Solutions. *Comput. Surveys* 56, 7 (2024), 1–29.
 - [130] Ryan Rossi and Nesreen Ahmed. 2015. The network data repository with interactive graph analytics and visualization. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 29.

- [131] Julian B Rotter. 1967. A new scale for the measurement of interpersonal trust. *Journal of personality* (1967).
- [132] Fatai Sado, Chu Kiong Loo, Wei Shiung Liew, Matthias Kerzel, and Stefan Wermter. 2023. Explainable goal-driven agents and robots-a comprehensive review. *Comput. Surveys* 55, 10 (2023), 1–41.
- [133] Rashmi P Sarode, Yutaka Watanobe, and Subhash Bhalla. 2023. A Decentralized Blockchain Powered Social Network for Secure and Transparent Online Interactions. In *Proceedings of the 2023 4th Asia Service Sciences and Software Engineering Conference*. 141–147.
- [134] Akрати Saxena, George Fletcher, and Mykola Pechenizkiy. 2024. Fairnsa: Algorithmic fairness in social network analysis. *Comput. Surveys* 56, 8 (2024), 1–45.
- [135] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. 2008. The graph neural network model. *IEEE transactions on neural networks* 20, 1 (2008), 61–80.
- [136] Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina Eliassi-Rad. 2008. Collective classification in network data. *AI magazine* 29, 3 (2008), 93–93.
- [137] Glenn Shafer. 1976. *A mathematical theory of evidence*. Vol. 42. Princeton university press.
- [138] Qusai Shambour and Jie Lu. 2015. An effective recommender system by unifying user and item trust information for B2B applications. *J. Comput. System Sci.* 81, 7 (2015), 1110–1126.
- [139] Wanita Sherchan, Surya Nepal, and Cecile Paris. 2013. A survey of trust in social networks. *ACM Computing Surveys (CSUR)* 45, 4 (2013), 1–33.
- [140] Shuhao Shi, Kai Qiao, Jian Chen, Shuai Yang, Jie Yang, Baojie Song, Linyuan Wang, and Bin Yan. 2023. Mgtab: A multi-relational graph-based twitter account detection benchmark. *arXiv preprint arXiv:2301.01123* (2023).
- [141] Xing Su, Jian Yang, Jia Wu, and Yuchen Zhang. 2023. Mining user-aware multi-relations for fake news detection in large scale online social networks. In *Proceedings of the sixteenth ACM international conference on web search and data mining*. 51–59.
- [142] Alistair G Sutcliffe, Di Wang, and Robin IM Dunbar. 2015. Modelling the role of trust in social relationships. *ACM Transactions on Internet Technology (TOIT)* 15, 4 (2015), 1–24.
- [143] Steve Tadelis. 2013. *Game theory: an introduction*. Princeton university press.
- [144] Jiliang Tang, Huiji Gao, and Huan Liu. 2012. mTrust: Discerning multi-faceted trust in a connected world. In *Proceedings of the fifth ACM international conference on Web search and data mining*. 93–102.
- [145] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. 2008. Arnetminer: extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. 990–998.
- [146] Lei Tang and Huan Liu. 2009. Relational learning via latent social dimensions. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. 817–826.
- [147] Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology* 29, 1 (2010), 24–54.
- [148] Takane Ueno, Yuto Sawa, Yeongdae Kim, Jacqueline Urakami, Hiroki Oura, and Katie Seaborn. 2022. Trust in human-AI interaction: Scoping out models, measures, and methods. In *CHI conference on human factors in computing systems extended abstracts*. 1–7.
- [149] Farman Ullah and Sungchang Lee. 2017. Community clustering based on trust modeling weighted by user interests in online social networks. *Chaos, Solitons & Fractals* 103 (2017), 194–204.
- [150] Muhammad Habib ur Rehman, Khaled Salah, Ernesto Damiani, and Davor Svetinovic. 2019. Trust in blockchain cryptocurrency ecosystem. *IEEE Transactions on Engineering Management* 67, 4 (2019), 1196–1212.
- [151] Raquel Urena, Gang Kou, Yucheng Dong, Francisco Chiclana, and Enrique Herrera-Viedma. 2019. A review on trust propagation and opinion dynamics in social networks and group decision making frameworks. *Information Sciences* 478 (2019), 461–475.
- [152] Nikhita Vedula, Srinivasan Parthasarathy, and Valerie L Shalin. 2017. Predicting trust relations within a social network: A case study on emergency response. In *Proceedings of the 2017 ACM on Web Science Conference*. 53–62.
- [153] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, Yoshua Bengio, et al. 2017. Graph attention networks. *stat* 1050, 20 (2017), 10–48550.
- [154] VG Vinod Vydiswaran, ChengXiang Zhai, and Dan Roth. 2011. Content-driven trust propagation framework. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. 974–982.
- [155] Kevin S Walsh, David P McGovern, Andy Clark, and Redmond G O’Connell. 2020. Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the new York Academy of Sciences* 1464, 1 (2020), 242–268.
- [156] Jingwen Wang, Xuyang Jing, Zheng Yan, Yulong Fu, Witold Pedrycz, and Laurence T Yang. 2020. A survey on trust evaluation based on machine learning. *ACM Computing Surveys (CSUR)* 53, 5 (2020), 1–36.
- [157] Jie Wang, Zheng Yan, Jiahe Lan, Elisa Bertino, and Witold Pedrycz. 2024. Trustguard: Gnn-based robust and explainable trust evaluation with dynamicity support. *IEEE Transactions on Dependable and Secure Computing* (2024).

- [158] Qi Wang, Weiliang Zhao, Jian Yang, Jia Wu, Shan Xue, Qianli Xing, and S Yu Philip. 2021. C-DeepTrust: A context-aware deep trust prediction model in online social networks. *IEEE Transactions on neural networks and learning systems* 34, 6 (2021), 2767–2780.
- [159] Qi Wang, Weiliang Zhao, Jian Yang, Jia Wu, Chuan Zhou, and Qianli Xing. 2020. Atne-trust: attributed trust network embedding for trust prediction in online social networks. In *2020 IEEE International Conference on Data Mining (ICDM)*. IEEE, 601–610.
- [160] Xiaoming Wang, Xinyan Wang, Geyong Min, Fei Hao, and CL Philip Chen. 2020. An efficient feedback control mechanism for positive/negative information spread in online social networks. *IEEE transactions on cybernetics* 52, 1 (2020), 87–100.
- [161] Yao Wang and Julita Vassileva. 2003. Bayesian network-based trust model. In *Proceedings IEEE/WIC International Conference on Web Intelligence (WI 2003)*. IEEE, 372–378.
- [162] Jie Wen, Nan Jiang, Lang Li, Jie Zhou, Yanpei Li, Hualin Zhan, Guang Kou, Weihao Gu, and Jiahui Zhao. 2024. TA-Detector: A GNN-based Anomaly Detector via Trust Relationship. *ACM Transactions on Multimedia Computing, Communications and Applications* (2024).
- [163] Jie Wen, Nan Jiang, Lang Li, Jie Zhou, Yanpei Li, Hualin Zhan, Guang Kou, Weihao Gu, and Jiahui Zhao. 2025. TA-detector: a GNN-based anomaly detector via trust relationship. *ACM Transactions on Multimedia Computing, Communications and Applications* 21, 8 (2025), 1–21.
- [164] Jian Wu, Mingshuo Cao, Francisco Chiclana, Yucheng Dong, and Enrique Herrera-Viedma. 2020. An optimal feedback model to prevent manipulation behavior in consensus under social network group decision making. *IEEE Transactions on Fuzzy Systems* 29, 7 (2020), 1750–1763.
- [165] Jian Wu, Sha Wang, Francisco Chiclana, and Enrique Herrera-Viedma. 2021. Two-fold personalized feedback mechanism for social network consensus by uninorm interval trust propagation. *IEEE Transactions on Cybernetics* 52, 10 (2021), 11081–11092.
- [166] Shiwen Wu, Fei Sun, Wentao Zhang, Xu Xie, and Bin Cui. 2022. Graph neural networks in recommender systems: a survey. *Comput. Surveys* 55, 5 (2022), 1–37.
- [167] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. 2020. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems* 32, 1 (2020), 4–24.
- [168] Yunpeng Xiao, Diqiang Chen, Shihong Wei, Qian Li, Haohan Wang, and Ming Xu. 2019. Rumor propagation dynamic model based on evolutionary game and anti-rumor. *Nonlinear Dynamics* 95 (2019), 523–539.
- [169] Rongwei Xu, Guanfeng Liu, Yan Wang, Xuyun Zhang, Kai Zheng, and Xiaofang Zhou. 2024. Adaptive Hypergraph Network for Trust Prediction. *arXiv preprint arXiv:2402.05154* (2024).
- [170] Rongwei Xu, Guanfeng Liu, Yan Wang, Xuyun Zhang, Kai Zheng, and Xiaofang Zhou. 2025. Intent-based Trust Evaluation. *IEEE Transactions on Knowledge and Data Engineering* (2025).
- [171] Hao Xue, Qiaozhi Wang, Bo Luo, Hyunjin Seo, and Fengjun Li. 2019. Content-aware trust propagation toward online review spam detection. *Journal of Data and Information Quality (JDIQ)* 11, 3 (2019), 1–31.
- [172] Ronald R Yager. 2008. Entropy and specificity in a mathematical theory of evidence. *Classic works of the Dempster-Shafer theory of belief functions* (2008), 291–310.
- [173] Zheng Yan, Li Peng, Wei Feng, and Laurence T Yang. 2021. Social-chain: Decentralized trust evaluation based on blockchain in pervasive social networking. *ACM Transactions on Internet Technology (TOIT)* 21, 1 (2021), 1–28.
- [174] Kai-Cheng Yang, Onur Varol, Pik-Mai Hui, and Filippo Menczer. 2020. Scalable and generalizable social bot detection through data selection. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34. 1096–1103.
- [175] Haikun Yu, Dacheng Jiang, Guipeng Zhang, Zhenguo Yang, and Wenyn Liu. 2023. A reputation system based on blockchain and deep learning in social networks. In *2023 26th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. IEEE, 630–635.
- [176] Su-Min Yu, Zhi-Jiao Du, Xue-Yang Zhang, Han-Yang Luo, and Xu-Dong Lin. 2021. Trust Cop-Kmeans clustering analysis and minimum-cost consensus model considering voluntary trust loss in social network large-scale decision-making. *IEEE Transactions on Fuzzy Systems* 30, 7 (2021), 2634–2648.
- [177] Lotfi A Zadeh. 1965. Fuzzy sets. *Information and Control* (1965).
- [178] Lotfi Asker Zadeh. 1975. The concept of a linguistic variable and its application to approximate reasoning—I. *Information sciences* 8, 3 (1975), 199–249.
- [179] Razieh Nokhbeh Zaeem, David Liao, and K Suzanne Barber. 2019. Predicting disease outbreaks using social media: Finding trustworthy users. In *Proceedings of the Future Technologies Conference (FTC) 2018: Volume 1*. Springer, 369–384.
- [180] Hangjing Zhang, Yuejiang Li, Yan Chen, and H Vicky Zhao. 2020. Smart evolution for information diffusion over social networks. *IEEE Transactions on Information Forensics and Security* 16 (2020), 1203–1217.
- [181] Hengjie Zhang, Fang Wang, Yucheng Dong, Francisco Chiclana, and Enrique Herrera-Viedma. 2021. Social trust driven consensus reaching model with a minimum adjustment feedback mechanism considering assessments-modifications willingness. *IEEE Transactions on Fuzzy Systems* 30, 6 (2021), 2019–2031.

- [182] Yizhou Zhang, Karishma Sharma, and Yan Liu. 2021. Vigdet: Knowledge informed neural temporal point process for coordination detection on social media. *Advances in Neural Information Processing Systems* 34 (2021), 3218–3231.
- [183] Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli, and Christopher D Manning. 2017. Position-aware attention and supervised data improve slot filling. In *Conference on empirical methods in natural language processing*.
- [184] Huan Zhao, Xiaogang Xu, Yangqiu Song, Dik Lun Lee, Zhao Chen, and Han Gao. 2018. Ranking users in social networks with higher-order structures. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.
- [185] Ke Zhou, Hongyuan Zha, and Le Song. 2013. Learning social infectivity in sparse low-rank networks using multi-dimensional hawkes processes. In *Artificial intelligence and statistics*. PMLR, 641–649.
- [186] Michelle X Zhou, Huahai Yang, Gloria Mark, Mengdie Hu, Jalal Mahumd, and Aditya Pal. 2021. Building personalized trust: discovering what makes one trust and act on Facebook posts. *ACM Transactions on Social Computing (TSC)* 4, 3 (2021), 1–28.
- [187] Xinyi Zhou and Reza Zafarani. 2020. A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)* 53, 5 (2020), 1–40.
- [188] Yang Zhou, Ling Liu, Kisung Lee, Balaji Palanisamy, and Qi Zhang. 2020. Improving collaborative filtering with social influence over heterogeneous information networks. *ACM Transactions on Internet Technology (TOIT)* 20, 4 (2020), 1–29.

Received 21 December 2025

