IRAQI
Academic Scientific Journals

Available at https://www.iasj.net/iasj
Iraqi Academic Scientific Journals
Journal homepage:
https://journals.uokerbala.edu.iq/index.php/UOKJ

مجلة جامعة كربلاء
Journal Of Kerbala University

Review Article

Heuristic Surveying Trust in Networks: Classical Models and Modern Deep Learning Approaches

[1,] Aseel Hussein Zahi [2,]Diyar M. Witefee
[1,2,]Ministry of Education, General Directorate of Education in Babylon, Babylon, Iraq

Article Info
Article history:
Received 29 -7-2025
Received in revised form 3-9-2025
Accepted27-10-2025
Available online 31 -12 -2025
Keywords: Network trust; trust evaluation; graph neural networks (GNNs); attention mechanisms; peer-to-peer (P2P) networks; trust representation learning; multimodal trust fusion.

Abstract:

Early trust frameworks relied on heuristic or graph-theoretic rules (e.g., EigenTrust, TrustRank) but often suffered from brittle propagation semantics and limited adaptability to dynamic behaviors. In recent years, data-driven approaches—particularly graph neural networks (GNNs), attention-based models, and hybrid deep learning frameworks—have catalyzed significant advances in trust evaluation by learning the latent representation of trust features and propagation patterns directly from data. This review presents a comprehensive survey of network trust research, encompassing foundational definition and taxonomies, classical computational methods, machine learning and deep learning innovations (with emphasis on GNNs and attention mechanisms), evaluation protocols and datasets, key applications (social recommender systems, IoT security, blockchain, and fraud detection), as well as ongoing challenges (data sparsity dynamic adaptation, explainability, and robustness). Network trust quantifies the confidence in relationships among entities (e.g., users, devices, organizations) in interconnected systems, ranging from social networks to the Internet of Things (IoT) and peer-to-peer (P2P) environments. We conclude by outlining promising research directions, such as self-supervised trust representation learning, lifelong adaptation, explainable trust frameworks, privacy-preserving trust computation, and multimodal trust fusion, to guide future work toward robust, scalable, and interpretable trust mechanisms. This review synthesizes the current state of the art in network trust, providing researchers and practitioners with a structured roadmap of methodologies, datasets, applications, and future research directions.

Corresponding Author E-mail: Aseel1371983@gmail.com, Dyarz2017@gmail.com

Peer review under responsibility of Iraqi Academic Scientific Journal and University of Kerbala.

1. Introduction

As digital interactions proliferate through social media, e-commerce, IoT devices, and decentralized recommendations, there is a need for resilience against malicious behaviors. In networked contexts, trust is typically defined as a measure of how much a *trustor* believes a *trustee* will behave according to certain expectations (e.g., honesty, reliability, integrity) based on past interactions, endorsements, or observed behaviors [1],[2].

Traditional trust models-such as reputation systems in P2P networks-laid the groundwork by aggregating local feedback and propagating trust metrics via graph traversals (e.g., eigentrust, TrustRank). However, these systems often struggled with adversarial manipulations (e.g., Sybil attacks), lacked adaptability to rapidly evolving networks, and depended heavily on preset heuristics instead of data-driven learning [3],[4].

The advent of machine learning, particularly deep learning, has revolutionized trust modeling by enabling end-to-end training of trust predictors from raw network data. Graph Neural Networks (GNNs), which leverage message passing and neighborhood aggregating, can capture complex relational patterns and trust propagation semantics. Meanwhile, attention-based architectures allow differential weighting of neighbor contributions, improving robustness to noisy or malicious links. These deep models have demonstrated superior performance on large-scale, real-world trust public trust datasets and open new avenues for dynamic, interpretable, and multi-source trust evaluation [3],[4].

Despite rapid progress, major challenges persist: (1) data scarcity and cold-start issues limit model training when trust relationships are sparse; (2) dynamic adaptation is required to update trust scores in real time as network topologies evolve; (3) explainability is critical in high-stakes contexts (e.g., finance, healthcare) but often lacking in black-box deep models; and (4) adversarial robustness remains an open problem, as trust networks are susceptible to collusion and fake endorsements. Addressing these challenges requires interdisciplinary efforts spanning graph representation learning, explainable AI, secure multi-party computation, and multimodal data fusion. This review synthesizes the current state of the art in network trust, providing researchers and practitioners with a structured roadmap of methodologies, datasets, applications, and future research directions. We surveyed computational trust in networks with emphasis on classical methods (heuristics, graph-theoretic, probabilistic) and learning-based models (ML, GNNs, Transformers).

2. Foundations of Network Trust

2.1 Definition and Taxonomies

Trust has been studied in sociology, psychology, and economics long before its formalization in network computing. In networked systems, trust typically refers to a quantitative or qualitative measure-often a real-valued score-reflecting a trustor's confidence in a trustee's reliability or integrity based on direct interactions, endorsements from others, or inferred behaviors from the network topology. “To clarify the landscape, Table 1 compares how different trust representations and model families perform along six practical dimensions: adaptability to dynamics, interpretability, robustness to adversarial noise, uncertainty handling, data requirements, and scalability.

Table 1: Trust Representation comparison

Category	Typical Examples	Adaptability to Dynamics	Interpretability	Robustness (Adversarial/Noise)	Uncertainty Handling	Data Requirement	Scalability
Binary trust	{0,1} decisions	Low	High	Low–Medium	Weak	Low	High
Continuous trust	score ∈ [0,1]	Medium	Medium	Medium	Medium–High	Medium	High
Heuristic / Rule-based	EigenTrust, TrustRank	Low–-Medium	High	Medium	Low	Low	High
Graph-theoretic	PageRank-like, flow/path	Medium	Medium	Medium	Low–Medium	Low–Medium	High
Probabilistic / Bayesian	Subjective Logic, Bayesian trust	Medium	Medium–High	Medium	High	Medium	Medium
Feature-based ML	LR, RF, SVM, XGBoost	Medium	Medium	Medium	Medium	Medium–High	High
Deep Learning (non-graph)	MLP/ CNN on features	Medium	Medium	Medium	Medium	High	Medium–High
GNN / Attention	TrustGNN, KGTrust, TrustGuard	High	Medium partially explainable)	Relatively High	Medium–High	High	Medium–High

Researchers classify trust along several dimensions:

- **Direct vs. Indirect Trust**:

○ Direct trust is computed from first-person experiences (e.g., user A rates user B positively based on a completed transaction).

- **Static vs. Dynamic Trust**:

○ Static trust assumes trust relationships are relatively stable over time, suitable for environments with infrequent interaction changes.

○ Indirect trust (or inferred trust) aggregates trust recommendations via intermediate nodes (e.g., A trusts C because A trusts B and B trusts C, subject to decay or confidence adjustments) [2].

○ Dynamic trust continuously updates trust scores as new interactions occur, accounting for temporal factors (e.g., recent misbehavior might rapidly decrease trust)[3].

○

○ **Binary vs. Continuous Trust**:

○ Binary models classify entities as "trustworthy" or "untrustworthy," often based on thresholded scores.

○ Continuous models assign a real-valued trust score (e.g., on a [0,1] scale), enabling finer-grained differentiation among trustees [2].

We categorize traditional trust models by abstraction level and methodology (Table 2): heuristic/rule-based [1],[2], graph-theoretic [1],[2], probabilistic/Bayesian [8], feature-based ML [8], and deep learning (GNN/attention) [5],[6].

Table 2: Traditional Trust Models

Method	Pros	Cons
Heuristic / Rule-based	Simple, transparent, fast	Fragile to change; easy to game
Graph-theoretic	Scalable; no labels needed	Popularity ≠ trust; topology bias
PageRank-like (EigenTrust/TrustRank)	Stable global scores; noise damping	Needs seeds; adapts slowly
Path / Flow aggregation	Captures transitivity; explainable paths	Path explosion; shortcut attacks
Local k-NN / Majority	Simple; tunable (k)	Weak in sparsity; echo chambers
Probabilistic / Bayesian	Uncertainty-aware; incremental	Prior-sensitive; heavier at scale
Subjective Logic	Rich belief/uncertainty semantics	Operator/parameter sensitive
Dempster–Shafer	Handles conflicting evidence	Expensive at high conflict
Fuzzy Logic	Human-readable; smooth vagueness	Subjective tuning; scalability varies
Matrix Factorization	Works with sparsity; latent traits	Cold-start;low interpretability
Game-theoretic / Incentive	Attack-aware by design	Needs accurate utility models
Temporal Filtering	Adapts quickly; denoises	May forget long-term evidence

Figure 1: provides a compact overview of the workflow: inputs , preprocessing , GNN model then evaluation metrics, which summarize how move from data to results.

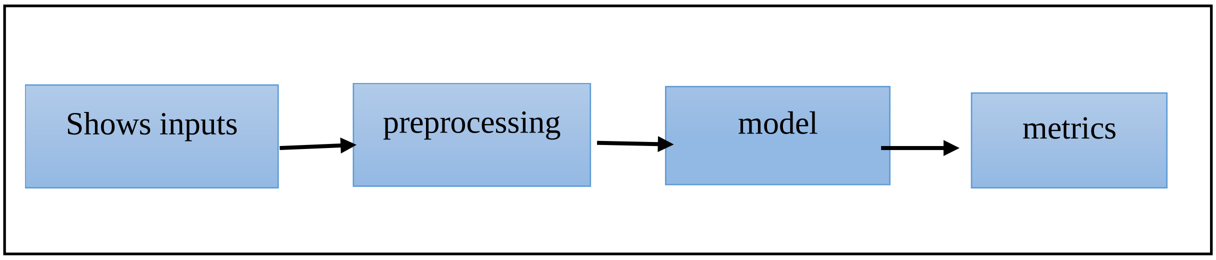

Figure 1: GNN workflow

2.2 Key Concepts

Several fundamental concepts underlie trust computation:

- **Trust propagating**: the mechanism by which trust scores spread through the network. Simple transitive rules (e.g., trust(A→C) = trust(A→B) × trust(B→C)) often lead to overestimation or amplification of errors [1]. Advanced methods introduce decay factors, probabilistic propagation, or energy-based flow to limit trust dilution over multiple hops [5].
- **Trust Aggregating**: combining multiple trust evidences, such as direct feedback, group endorsements, and feature-based scores, into a single consolidated trust value. Heuristic methods often employ weight averaging or maximum selectors, whereas data-driven models learn aggregating functions (e.g., attention weights in GNNs) that dynamically adjust to the context. For example, attention-based trust models assign higher weights to neighbors with stronger reputations or more relevant features, improving robustness against noisy edges [7].
- **Trust Dynamics:** Because trust relationships evolve, especially in highly active network models, models must account for temporality. Techniques include temporal decay functions (older interactions contribute less), sliding windows of recent behaviors, or explicit recurrent models (e.g., temporal GNNs) to capture evolving trust patterns. Trust Guard, for instance, uses a temporal attention layer to capture sequence dependencies in trust evolution while defending against temporal attacks [8].

3. Traditional Trust Computation Methods

Before the rise of data-driven learning, trust computation primarily relied on heuristics, graph flows, and statistical reasoning.

Below, we summarize key classical approaches.

3.1 Heuristic and Reputation Systems

Eigen Trust [1] (Kamvar, Schlosser, & Garcia-Molina, 2003) computes global reputation scores in P2P networks by normalizing local trust ratings (i.e., signed counts of satisfactory or unsatisfactory transactions) and iteratively propagating them via power iteration, analogous to PageRank-until convergence. Each peer aggregates normalized feedback from neighbors, forming a probability distribution over trustees; repeated propagation yields global trust vectors that mitigate malicious peers through global consensus. Eigen Trust effectively reduces inauthentic file sharing (e.g., free-riding) but assumes an honest majority and requires a seed set of pre-trusted nodes for stability.

TrustRank [2] (Gyöngyi, Garcia-Molina, & Pedersen, 2004) was developed to combat web spam by semi-automatically selecting a small set of reputable "seed" pages labeled by experts. Trust is then propagated to neighboring pages via a modified PageRank process with a high teleportation probability back to the seed set. Pages closer in link distance to trustworthy seeds inherit higher trust, while distant pages accumulate less. TrustRank significantly reduces link-spam influence but depends on a carefully chosen seed set to avoid bias and requires periodic reselection as the web evolves.

CredibleRank and Anti-TrustRank extend TrustRank's framework by integrating topical relevance (e.g., Topical TrustRank uses topic-sensitive teleportation) or propagating distrust from known spam sources. These extensions aim to improve spam detection in specialized domains (e.g., political blogs, e-commerce reviews) but still require manual seed curation and struggle with dynamically changing link patterns.

3.2 Graph-based Probabilistic Models

Probabilistic trust models treat trust as a random variable with a probability distribution. Bayesian Trust Models use Bayes's theorem to update trust beliefs upon observing new interactions:

Here, priors $P(t_{A\rightarrow B})$ may derive from initial ratings or domain knowledge, while P(evidence|trust) models likelihoods of observed behaviors (e.g., positive/negative feedback counts). Though Bayesian models quantify uncertainty explicitly, they require careful prior selection and can be computationally intensive for large graphs-subjective logic extends Bayesian trust by representing trust opinions as triplets (b,d,u) (b,d,u)(b,d,u) denoting belief, disbelief, and uncertainty, respectively, with operators for combining opinions, but its complexity limits real-time scaling.

4. Machine Learning and Deep learning Approaches

Data-driven methods have transformed trust computation by automating feature extraction and modeling complex propagation patterns. We categorize these into classic machine learning, deep neural networks (DNNs), and Graph Neural Networks (GNNs) with attention[8].

4.1 Classic Machine Learning Models

In classic machine learning (ML) approaches, trust evaluation is framed as either a binary classification problem (trust versus distrust) or a regression problem (predicting continuous trust scores). Input features typically include:

- interaction-based features, such as the number of positive/negative transactions, the recency of the last interaction, and the number of mutual connections;
- Features of a textual nature, such as sentiment polarity or topic distributions, can be found in user reviews, comments, or emails.
- Network structural features, such as local clustering coefficients, shortest path lengths, Katz scores, and node centrality measures.

Models such as logistic regression, decision trees, random forests, and gradient boosting (e.g., XGBoost) learn to predict trust based on historical, labeled relations. Unsupervised methods, such as clustering or anomaly detection, identify outliers or anomalous subgraphs (e.g. dense clusters of nodes with low trust who endorse each other) to flag potential collusion or denial-of-service attacks.

Although classic ML offers more flexibility than heuristics, it has limitations.

1. Manual feature engineering: Handcrafting relevant features is time-consuming and domain-specific.
2. These models cannot inherently propagate trust signals beyond immediate neighborhoods (i.e. they lack inductive graph learning capability).
3. Scalability constraints: Constructing large feature matrices and training on massive graphs can be expensive.

4.2 Deep Neural Networks (DNNs)

Deep learning methods initially applied feed-forward neural networks or CNNs of transformed feature vectors representing user pairs (e.g., concatenated user profiles, interaction histories, and network metrics). By stacking nonlinear layers, DNNs capture complex interactions among features, often outperforming shallow classifiers. For instance, early work in e-commerce trust prediction used DNNs to learn latent patterns from combined rating histories and user demographics. However, DNNs ignore the explicit graph topology, treating each pair independently and failing to leverage neighborhood information beyond engineered features [8].

4.3 Graph Neural Networks and Attention-Based Models

Graph Neural Networks (GNNs) [5] unify propagation and representation learning by operating directly on graph structures. Trust evaluation models often incorporate edge asymmetry (trust is rarely symmetric) and multi-hop propagation patterns into their aggregation functions. Notable GNN-based trust models are summarized in recent deep trust models: attention-based GNNs [6], TrustGNN [3], KGTrust [4], TrustGuard [5], and Transformer/Hybrid approaches [5] in Table 3, comparing datasets, performance, robustness, and computation.

Table 3: deep trust model comparison datasets, performance, robustness, and computation.

Model	Typical datasets	Reported gains (vs. traditional)	Robustness under attack	Params / Compute	Notes (what drives performance)
Attention-based NN	Epinions, CiaoDVD	↑ AUC-ROC/PR, ↑ P@K (moderate)	Medium (filters noisy neighbors)	Medium	Learns to weight neighbors; depends on feature quality.
TrustGNN	Epinions, CiaoDVD	↑↑ AUC-ROC/PR, ↑ P@K (consistent)	Medium–High (multi-hop smoothing)	Medium–High	Message passing captures propagation/asymmetry; strong all-rounder.
KGTrust	SIoT / hetero graphs	↑ AUC-ROC/PR in sparse/noisy links	Medium–High (semantic context helps)	High	Adds knowledge-graph semantics; needs curated KG and memory.
TrustGuard	Epinions (temporal), others	↑↑ AUC-PR/P@K esp. over time	High (spatial defense + temporal attention)	High	Robust to injected/malicious edges; higher tuning & training cost.
Transformer / Hybrid	Custom/large graphs	↑ AUC-PR on long-range deps	Medium (depends on setup)	High–Very High	Captures global context; quadratic attention can limit scale.

4.4 Transformer-based and Hybrid Trust Models

Transformers push attention beyond a node’s immediate neighbors, letting the model learn long-range dependencies that classical methods and shallow GNNs often miss. In trust graphs, this means they can naturally capture multi-hop endorsement chains, coordinated or collusive patterns, and cross-community signals—without hand-crafted propagation rules. Two main design lines have emerged. The first is graph-Transformers, which replace or augment message passing with global self-attention applied to node/edge embeddings (often using sparse attention or structural/positional encodings such as Laplacian features or relative positions). The second is hybrid models that keep local GNN aggregation but add global Transformer blocks for mixing (e.g., GNN→Transformer or Transformer→GNN). For temporal trust, sequence-aware variants (time-decay kernels, time positional codes) track how trust evolves and reduce leakage in snapshot datasets. In multimodal cases (text reviews + graph), cross-modal attention fuses textual evidence with structure to down-weight noisy neighbors and popularity bias.

5. Evaluation Frameworks and Datasets

5.1 Dataset Characteristics & Diagnostics

Our datasets are sparse, snapshot-based, and often positively imbalanced; in some cases, “trust” is only a proxy derived from interactions. These factors can bias results toward popular nodes and overstate performance on the majority class. We address this by reporting PR-AUC with confidence intervals, using time-aware or stratified splits, sanitizing graphs, documenting label mappings, and keeping proxy-trust results separate from explicit-trust datasets. Table 4 summarizes the structural properties of the trust graphs we use: size, sparsity, temporal dynamics, and known noise/quirks, because these factors strongly affect model design.

Overview. Let nnn be the number of users (nodes) and mm the number of directed trust edges. Graph density $\delta=\frac{m}{n(n-1)}$ (no self-loops) quantifies sparsity.

Table 4: Structural properties of the trust graphs

Dataset	Node (n)	Edges (m)	Density (δ)	Timestamps	Trust label type	Sparsity notes	Annotation reliability / known noise
Epinions (signed who-trusts-whom)	131,828	841,372	4.84×10^{-5}	Snapshot (2003)	Explicit (trust & distrust)	Highly sparse,	Some releases contain self-loops; snapshot only (risk of temporal leakage).
Epinions (unsigned, soc-Epinions1)	75,879	508,837	8.84×10^{-5}	Snapshot (2003)	Explicit (trust only)	Highly sparse	Unsigned variant; often used for link prediction; snapshot only.
CiaoDVD (trust)	4,658	40,133	1.85×10^{-3}	Year tags (often treated static)	Explicit	Sparse, but denser than Epinions	Moderate size; connectivity better than Epinions.
Advogato (developer trust)	6,541	51,127	1.20×10^{-3}	Multiple snapshots (often static)	Explicit (multi-level)	Sparse	Weighted, directed; level mapping choices can affect labels.
PGP Web of Trust	14,367 → 31,524	37,900 →168,559	$\sim1.8\times10^{-4} \rightarrow 1.7\times10^{-4}$	Strong temporal signal	Explicit (cert/signature)	Sparse, evolving	Good for temporal studies; care with versioning/duplicates.
Twitter follower graph (proxy)	41,652,230	1,468,364,884	8.46×10^{-7}	Clear temporal evolution	Implicit / proxy (follow/RT)	Extremely sparse at scale	Proxy ≠ trust: popularity bias; must flag label provenance.

5.2 Evaluation Metrics

In practice, trust models are judged on predictive performance, ranking quality, robustness, efficiency, and interpretability—plus domain fit in high-stakes settings [5].

- Regression metrics. For continuous trust scores, report MSE, MAE, and RMSE to quantify error against ground truth.
- Classification metrics. When trust is binarized (trust vs. distrust), use precision, recall, F1-score, and accuracy.
- Ranking metrics. Use AUROC, AUPRC, and Precision to check whether highly trusted nodes are ranked above untrusted ones.
- Robustness metrics. Simulate Sybil/collusion attacks (e.g., injecting malicious edges) and track trust degradation, false-positive rates, or attack success rates.
- Computational efficiency. Report training/inference time and memory footprint, which matter on graphs with millions of nodes and edges.
- Explain ability and interpretability. Combine user studies and model probes (e.g., inspecting attention weights in GNNs) to assess transparency. In high-stakes domains (healthcare, finance), expert review verifies that aggregated trust aligns with domain knowledge. In SIoT scenarios, for example, experts check whether device trust scores reflect realistic reliability and security requirements.

6. Applications of Network Trust

6.1 Social Recommender Systems

Trust-aware recommenders use user–user trust to weight neighbor contributions when predicting preferences. In collaborative filtering (CF), the user–item matrix is augmented with trust links; propagation over these links helps with sparsity and cold-start by borrowing signal from indirectly trusted neighbors [9].

- Trust-aware matrix factorization. Add a trust regularizer that discourages large divergences between the latent factors of trusted users, improving rating prediction when explicit ratings are scarce.
- GNN-based trust recommenders (e.g., GraphRec, TrustGNN). Learn joint user–item embeddings by propagating trust signals alongside rating interactions, often outperforming classic baselines (e.g., BPR, TrustSVD) on Epinions and CiaoDVD.
- Hybrid attention models. Fuse review-text sentiment (from CNNs/Transformers) with structural trust embeddings via multi-head attention to produce more personalized rankings.

Empirical pattern. Compared with rating-only models, trust-based systems typically achieve higher Precision@K and lower RMSE, especially under high sparsity (e.g., <5% observed ratings). Caveats. They depend on reliable trust data, and capturing temporal trust dynamics remains challenging as social ties evolve. [9]

6.2 Security and Privacy in IoT

In IoT networks, heterogeneous devices comprise [4](e.g., sensors, actuators, gateway)- trust computation safeguards data integrity, access control, and secure communication. Use cases involve:

- **Malicious Node Detection**: SIoT trust models (e.g., KGTrust) integrate device behavior logs (e.g., abnormal data spikes, missing heartbeats) and external semantic knowledge (e.g., device metadata) into GNNs to predict trustworthiness. Enabling early blacklisting of compromised devices.
- **Secure Routing**: trust-based routing protocols assign trust paths to avoid malicious or unreliable nodes in multi-hop sensor networks. Dynamic trust evaluation helps maintain the quality of service under node failures or jamming attacks.
- **Access Control and Authentication**: trust scores feed into distributed access control mechanisms. Devices with trust above a threshold gain privileges, while low-trust nodes require additional authentication or are quarantined.

Challenges include resource constraints (computational and energy limitations), dynamic topology changes (mobile IoT devices), and privacy concerns (raw interaction logs may contain sensitive information). Privacy-preserving trust (via federated learning or homomorphic encryption) aims to compute trust scores collaboratively without exposing local data.

6.3 Blockchain and Decentralized Systems

In decentralized ledgers and peer-to-peer energy trading platforms, trust metrics complement consensus mechanisms to improve network efficiency and security [10]:

- **Permissioned Blockchains**: Use trust scores to rank validator nodes in delegated proof-of-stake (DPoS) systems, ensuring that highly reputable nodes participate in block validation.
- **Decentralized Identity (DID)**: Trust evaluation frameworks feed into DID systems, where trust scores derived from on-chain transaction history and off-chain reputation (e.g., user KYC documents) inform authentication and access decisions.
- **Cross-Chain Interoperability**: Trust metrics assist in selecting reliable gateway relayers for asset transfers, mitigating assets' loss due to malicious or underperforming relayers.

While consensus algorithms enforce data. Integrity, trust scores provide an additional layer of security by identifying colluding or dishonest nodes. Integrating trust with blockchain requires careful design to avoid double-counting (trust recorded on-chain could be manipulated) and maintain low transaction overhead.

6.4 Fraud Detection and Cybersecurity

Trust evaluation plays a pivotal role in identifying fraudulent users and malicious behaviors in financial networks, e-commerce platforms, and online marketplaces[5]:

- **Sybil and Bot Detection**: GNN-based trust models detect Sybil nodes (multiple fake accounts controlled by one adversary) by identifying anomalous trust propagation patterns (e.g., densely interconnected low-trust clusters). Robust aggregation methods (e.g., TrustGuard's spatial defense) mitigate Sybil infiltration by down-weighting suspicious edges.
- **Transaction Fraud:** In payment networks, trust scores guide risk assessment. If a user's trust falls below a threshold, additional verification (e.g., two-factor authentication) is required before processing high-value transactions.
- **Fake Review and Review Spam:** Trust-based reputation systems weight user reviews according to reviewer trust, reducing the influence of fake or manipulated reviews. Attention-based GNNs can detect colluding reviewers by spotting tight clusters of mutually endorsing untrustworthy accounts.

Empirical studies demonstrate that incorporating trust features reduces false-positive rates in fraud detection by 15–30% compared to transaction-only models, while maintaining high recall. Yet, attackers continuously adapt (e.g., by creating synthetic trust relationships), underscoring the need for robust and adaptive trust frameworks.

7. Challenges and Open Issues

Despite significant advances in trust modeling, several critical challenges remain [5],[6]:

1. Data Sparsity and Cold-Start: Many networked systems have extensive explicit trust labels; new users or devices have no interaction history, making trust estimation difficult. Self-supervised learning (e.g., contrastive objectives on graph structure) can alleviate this by leveraging unlabeled data, but performance still lags when interactions are extremely scarce reseachgate.netresearchgate.net
2. Dynamic and Evolving Networks: In real-world environments (e.g., social media, IoT), trust relationships change rapidly. Retraining GNNs from scratch upon each network update is infeasible. Online or incremental GNN architectures (e.g.,

temporal GNNs with memory modules) are necessary to support continuous learning, but designing stable update mechanisms that avoid catastrophic forgetting remains open.

3. **Explainability & interpretability.** Deep models—especially graph neural networks (GNNs)—are often seen as black boxes. In high-stakes settings (e.g., healthcare IoT, financial services), stakeholders need transparent justifications for trust scores. Attention visualizations can show which neighbors influenced a decision, but comprehensive frameworks for explanation (e.g., counterfactual analyses, rule extraction) are still limited.
4. **Adversarial attacks & robustness.** Malicious actors can game trust propagation by adding bogus edges or forming collusive cliques. Robust aggregation (e.g., defense layers, as in TrustGuard) and adversarial training mitigate some risks; however, strong, provable guarantees against adaptive Sybil or camouflage attacks are rare. Establishing theoretical robustness bounds under explicit threat models remains an open question.
5. **Heterogeneity & multimodality.** Modern networks blend text, images, geolocation, and device telemetry with graph structure. Integrating these heterogeneous signals into a unified trust model—via multimodal GNNs or cross-modal attention—poses architectural and optimization challenges, especially when modalities differ in scale, noise, and missingness.
6. **Privacy-preserving trust computation.** Trust estimation often touches sensitive data (e.g., communication logs, transactions). Techniques such as federated learning, secure multi-party computation (SMPC), and homomorphic encryption improve privacy but introduce computational and communication overhead. Designing scalable, efficient, privacy-aware trust protocols for large graphs is still unresolved.

Progress will require cross-disciplinary collaboration across graph learning, cryptography, human–computer interaction, and network security.

8. Feature Research Directions

Based on the challenges identified, we propose several promising research directions:

1. **Self-supervised and Contrastive Learning for Trust:** Leveraging unlabeled graph data via contrastive objectives, such as maximizing agreement between connected nodes and minimizing it for disconnected ones, can alleviate data scarcity. Graph contrastive learning methods (e.g., GraphCL)can produce robust trust embeddings without explicit labels; integrating these into trust evaluation frameworks may improve cold-start performance [3].
2. **Continual and Lifelong Learning:** Developing incremental GNNs that update trust embeddings in real time as new edges appear, without retraining from scratch, is essential for evolving networks. Memory-augmented architectures (e.g., Graph Memory Networks) or dynamic graph transformers could provide stable performance while accommodating frequent topology changes [6].
3. **Explainable Trust Models**: Beyond visualizing attention weights, trust models should produce human-readable rationales—e.g., "User A trusts User B because of 3 high-quality endorsements from mutual connections C, D, and E." Techniques such as rule extraction, counterfactual explanations, or concept bottlenecks can help translate black-box predictions into interpretable narratives. Research on combining symbolic reasoning with neural trust models appears promising [4].
4. **Robust trust under adversarial threats.** Reliability in hostile environments requires formal threat models, specifying a Sybil attacker's budget, capabilities, and objectives, and provable robustness guarantees for trust algorithms. Adapting certified robustness methods from computer vision to GNN-based trust frameworks could provide theoretical assurances against malicious manipulations[10].

5. **Privacy-preserving trust computation.** Combining federated learning with secure aggregation enables decentralized training without sharing raw data**.** Differential privacy can inject noise into local gradients to protect sensitive information while retaining utility. Future work should balance privacy, accuracy, and communication efficiency at a large scale in distributed settings [10].
6. **Multimodal trust fusion.** Next-generation models should integrate heterogeneous signals like text sentiment**,** image features**,** geolocation traces**,** and graph topology into a coherent embedding. Cross-modal attention and multimodal Transformers can contextually focus on the most informative modality, improving trust estimation in domains such as e-commerce (images, reviews, interaction networks) and autonomous-vehicle networks (sensor telemetry, maps) [10].
7. **Standard benchmarks and open datasets.** The field needs large-scale, heterogeneous trust datasets that include timestamped relations, content features, and external knowledge, alongside standardized evaluation protocols (clear train/validation/test splits, attack-injection scenarios) [7].
8. **Beyond dyadic trust.** Most approaches model pairwise (dyadic) trust, yet real-world scenarios involve group trust (e.g., the reliability of a device consortium) and contextual trust (a seller may be trustworthy in electronics but not apparel). Hypergraph neural networks and other higher-order models can capture group-level dynamics and cross-domain trust shifts [8]

9. Conclusion

Network trust remains a cornerstone for secure, reliable, and personalized interactions in interconnected digital ecosystems. From early heuristic systems (e.g., EigenTrust, TrustRank) to cutting-edge GNN-based frameworks (e.g., TrustGNN, KGTrust, TrustGuard), researchers have made significant strides in modeling trust propagation, handling dynamic networks, and integrating heterogeneous data. Deep learning—particularly GNNs and attention-based architectures—has unlocked the ability to learn latent trust representations directly from data, improving predictive accuracy and resilience to noisy or malicious inputs.

Yet, critical challenges persist: data sparsity, dynamic adaptation, explainability, adversarial robustness, heterogeneity, and privacy concerns demand innovative solutions. Future research should emphasize self-supervised learning, continual adaptation, explainability, privacy-preserving protocols, and multimodal trust fusion, underpinned by standardized benchmarks and open datasets. By embracing these directions, the research community can build trustworthy, scalable, and transparent trust mechanisms—laying the foundation for secure and reliable digital interactions across social networks, loT, blockchain, and beyond.

References

1. S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina, "The EigenTrust algorithm for reputation management in P2P networks," in *Proc. 12th Int. Conf. World Wide Web (WWW '03)*, 2003, pp. 640–651. doi: 10.1145/775152.775242.
2. Z. Gyöngyi, H. Garcia-Molina, and J. Pedersen, "Combating web spam with TrustRank," in *Proc. 30th Int. Conf. Very Large Data Bases (VLDB '04)*, 2004.
3. C. Huo, D. Jin, C. Liang, D. He, T. Qiu, and L. Wu, "TrustGNN: Group neural network-based trust evaluation via learnable propagative and composable nature," *arXiv:2205.12784*, 2022.
4. Z. Yu, D. Jin, C. Huo, Z. Wang, X. Liu, H. Qi, J. Wu, and L. Wu, "KGTrust: Evaluating trustworthiness of SIoT via knowledge-enhanced graph neural networks," in *Proc. Web Conf. 2023 (WWW '23)*, 2023, pp. 727–736.
5. J. Wang, Z. Yan, J. Lan, E. Bertion, and W. Pedrycz, "TrustGuard: GNN-based robust and explainable trust evaluation with dynamicity support," *arXiv:2306.13339*, 2023.

6. Y. Xu, Z. Feng, X. Zhou, M. Xing, H. Wu, X. Xu, S. Chen, C. Wang, and L. Qi, "Attention-based neural networks for trust evaluation in online social networks," *Information Sciences*, vol. 605, pp. 507–522, 2023. doi: 10.1016/j.ins.2023.02.045.
7. W. Guo, B. Yang, and Y. Zhang, "Deep learning-based trust evaluation: A survey," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 1–27, 2022.
8. P. Agrawal and V. Singh, "Social networks data analytical approaches for trust-based recommender systems," *Data & Knowledge Engineering*, vol. 145, art. 101801, 2023. doi: 10.1016/j.datak.2023.101801.
9. Z. Li, W. Fang, C. Zhu, W. Chen, T. Hao, and W. Zhang, "Trust evaluation with deep learning in online social networks: A state-of-the-art review," in *Advanced Intelligent Computing Technology and Applications* (LNCS, vol. 14470). Cham, Switzerland: Springer, 2024, pp. 3–12.
10. E. Dai, T. Zhao, H. Zhu, J. Xu, Z. Guo, H. Liu, J. Tang, and S. Wang, "A comprehensive survey on trustworthy graph neural networks: Privacy, robustness, fairness, and explainability," *arXiv:2204.08570*, 2022.
11. B. Zhao, P. Xiong, and Q. Zhou, "Trust analysis improvement through deep learning for signed social networks," *Journal of Network and Computer Applications*, vol. 160, 2024.
12. A. Said and A. Al-Suwailem, "Exploring trust dynamics in online social networks: A social network analysis perspective," *Algorithms*, vol. 29, no. 3, p. 37, 2023.
13. S. Karim, G. Tong, J. Li, A. Qadir, U. Farooq, and Y. Yu, "Current advances and future perspectives of image fusion: A comprehensive review," *Information Fusion*, vol. 90, pp. 185–217, 2023.