

DataTrustGraph: Graph Learning for Dynamic Reliability Assessment of Enterprise Data

Sai Mohan Dasari

Submitted:01/08/2023

Revised:24/09/2023

Accepted:02/10/2023

Abstract: The trustworthiness of any given data record is becoming an integral part of the trustworthiness of the entities to which it is linked; as a result, the interconnected and evolving nature of data has become increasingly important to enterprise organizations. Traditional data quality assessment methods that are based on the periodic evaluation of data records, and that usually assess data quality at the record level, are not well suited to environments such as those where data is constantly ingested, records are dependent on each other across multiple systems, and adversarial manipulation is a possibility. This paper presents a synthesis of literature from graph anomaly detection, static and dynamic graph representation learning, knowledge graph embedding, data provenance management, and classical data quality theory, and proposes DataTrustGraph, which is a conceptual framework for continuous, graph-based, enterprise data reliability assessment. The framework represents enterprise entities, transformations and provenance events as a heterogeneous, evolving graph, on which reliability scores are propagated via attention-weighted neighborhood aggregation, similar to the propagation methods developed in graph attention and temporal graph networks. The synthesis is expressed as a composite reliability equation combining intrinsic, dimension-based data quality scoring with dynamically propagated neighborhood trust. The reviewed literature indicates that memory-based, inductive, and attention-based graph architectures are complementary in capturing the structural, semantic, and temporal aspects of data unreliability, and that the study of data provenance and data governance emphasizes the role of auditability and access control as essential, but often under-integrated, elements of trust computation. The paper concludes by identifying three directions for future research: incremental graph learning, explainable data quality, and agreed evaluation protocols.

Keywords: *graph neural networks; data quality; dynamic graphs; enterprise data governance; data provenance; trust propagation; anomaly detection; knowledge graph embedding; graph attention networks*

1. Introduction

The reliability of any particular data point is dependent on the reliability of the data that is connected to it, and that is dependent on the provenance, quality, and behavior of the connected data. Graph-based anomaly detection algorithms are considered to be effective techniques for detecting anomalies in these interconnected systems, and recently their applications have been reported in various fields such as finance, network security and social platforms, and detection algorithms are classified based on whether an anomaly occurs at node, edge, subgraph or graph snapshot level (Akoglu et al., 2015).

Past solutions for trust and reputation in distributed data systems have relied on decentralized
Principal Data / AI Engineer

ledger and blockchain technologies where distributed consensus and cryptographic verification replace a centralized authority. A comprehensive survey of such approaches catalogs taxonomies for trust computation, reputation propagation, and resistance to identity-forging attacks (Bellini et al., 2020). These taxonomies provide conceptual precedents for mechanisms that propagate reliability across enterprise data platforms spanning multiple business units and systems of record.

Empirical studies have illustrated that errors in the data, including missing values and label noise, as well as inconsistent formatting, can propagate through the model pipeline and negatively affect the accuracy of downstream machine learning tasks. This evidence points to the need for continuous, automated reliability assessment mechanisms that are embedded directly into

enterprise data systems, rather than applied as a separate pre-processing step before analysis or operation (Budach et al., 2022). Motivated by these factors, the present paper synthesizes the literature from graph representation learning, dynamic graph modeling, data provenance, and data quality theory to state DataTrustGraph, a conceptual framework where reliability is not regarded as an attribute of a record, but rather as a property of a graph that is dynamically propagated. The rest of the paper is

organized as follows: Section 2 reviews basic concepts for enterprise data quality and trust; Section 3 reviews the existing techniques for static graph representation learning; Section 4 reviews the techniques for dynamic and temporal graph learning; Section 5 presents the proposed framework illustrated in Figure 1; Section 6 formulates the composite reliability score; Section 7 discusses the implications for governance and open challenges; and Section 8 concludes the paper.

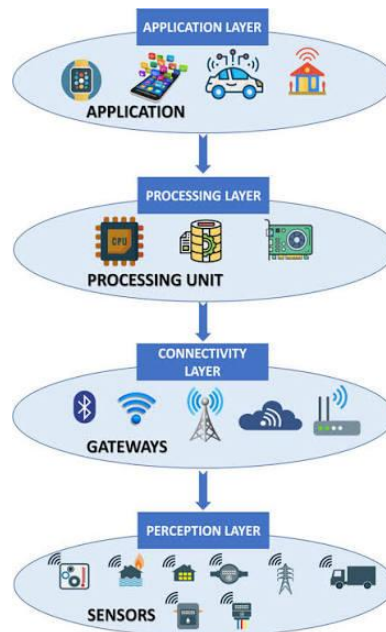


Figure 1

Four-Layer Architecture of the Proposed DataTrustGraph Framework (ResearchGate,2017)

2. Background: Enterprise Data Quality and Trust Foundations

These issues of measuring data quality at large scale have been explored in depth in the world of big data, where the volume, velocity, and variety of the data make rule-based data validation difficult. The field of work presented here has called for context-based, adaptive assessment methods that can handle data that is heterogeneous and rapidly changing, as opposed to data that meets a narrow definition and changes slowly, for which a fixed set of rules calibrated to standardize the data could be used (Cai & Zhu, 2015).

For financial fraud detection, in particular, the graph neural network architectures have been discussed as a growing method where transactions are represented as a graph, and fraudulent patterns are learned by performing iterative message propagation between entities connected by transactions. These reviews have emphasized that

these techniques can be applied to other enterprise reliability assessments beyond fraud detection, and that the principle used (trustworthiness of an entity is inferred from the trustworthiness and behavior of other entities) is easily extensible to other types of data pipelines (Cheng et al.,).

There have been many data quality frameworks proposed in both academic and industrial settings, and a systematic review of these frameworks has classified them into categories based on dimensions of data quality (accuracy, completeness, consistency and timeliness), and highlighted that there is significant variety in the way the dimensions are understood and measured in practice. This difference was identified as a challenge towards the integrative, cross-system reliability assessment, as a "good" record may be found adequate in one organization, but insufficient in another, and a shared graph-native representation, that can harmonize heterogeneous quality signals, is needed (Cichy & Rass, 2019). Figure 2 presents the

reviewed literature in chronological order, highlighting the shift from static to dynamic

architectures and then to provenance-aware architectures.

Table 1
Data Quality and Trust Foundations Synthesized from the Reviewed Literature

Reference	Focus Area	Core Contribution	Relevance to DataTrustGraph
Bellini et al. (2020)	Distributed trust systems	Taxonomy of blockchain-based trust and reputation computation	Precedent for decentralized, propagation-based trust scoring
Budach et al. (2022)	ML data quality impact	Empirical link between data defects and degraded model performance	Motivates continuous, embedded reliability monitoring
Cai & Zhu (2015)	Big data quality assessment	Argues for adaptive, context-aware quality frameworks	Basis for dynamic, non-static assessment logic
Cichy & Rass (2019)	Data quality frameworks	Comparative synthesis of accuracy, completeness, consistency, timeliness dimensions	Source of intrinsic node-level quality features

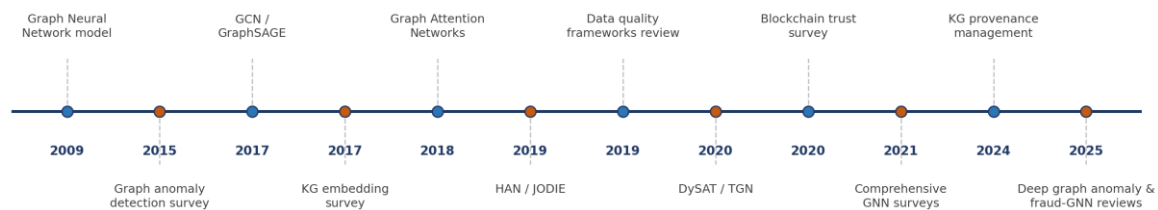


Figure 2
Chronological Progression of the Graph Learning and Data Quality Literature Reviewed

3. Foundations of Graph Representation Learning

Earlier transductive approaches are unable to generalize to nodes not seen during training; this limitation has been addressed by inductive representation learning. In this spirit, a popular approach is to learn node embeddings by using learned aggregator functions on sampled local neighborhoods, which allows representation learning on graphs that continually evolve as new entities are added after training (Hamilton et al., 2017).

Through a symmetrically normalized adjacency operator and a propagation rule per layer, spectral approximations of graph convolution operations have been formalized. The propagation rule is written as in Equation 1, where H is the matrix of node feature representations in a given layer, $\tilde{A}$ is the adjacency matrix with added self-loops, $\tilde{D}$ is the corresponding degree matrix, W is a trainable layer-specific weight matrix and σ is a nonlinear activation function, such as rectified linear unit.

$$H^{(l+1)} = \sigma \left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right) \quad (1)$$

This formulation has been proved to be useful for efficient semi-supervised classification on graph-structured data with a very small number of labeled nodes, and is a computationally tractable way to generalize spectral graph theory (Kipf & Welling, 2017). In the proposed model, a similar propagation rule is used, but instead of class labels,

it is used to propagate reliability indicators, e.g. the reliability of a data entity is derived from the reliability of the pipelines, tables and users with which it is structurally related. This neighborhood aggregation process is illustrated schematically in Figure 3.

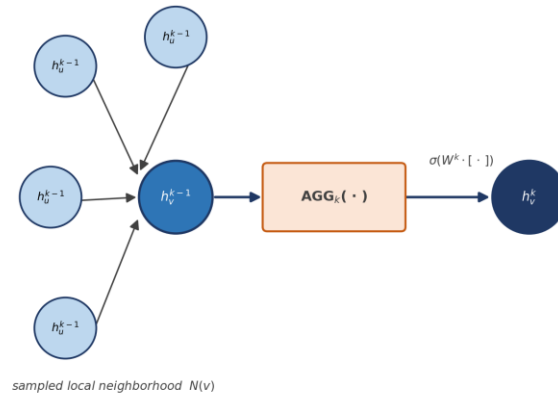


Figure 3

Inductive Neighborhood Aggregation Underlying Convolutional Reliability Propagation

Provenance management in knowledge graph platforms has been recognized as a prerequisite for building data trustworthiness. Proposed architectures include the inclusion of lineage metadata (such as data source, transformation history and responsible agent) as first-class graph entities that allow for the query and reasoning over it with the same traversal mechanisms as with substantive data (Kleinstauber et al.). Generally, an aggregation function for both

entity and provenance nodes that determines the representation of a node v at layer k , is formulated as follows (see Equation 2): the representation of v at layer k is calculated by concatenation of the representation of v at the previous layer and the aggregation of the prior-layer representations of its neighbors. A representative heterogeneous enterprise schema in which these provenance nodes are co-located with substantive entities is shown in Figure 4.

$$h_v^k = \sigma \left(W^k \cdot \text{CONCAT} \left(h_v^{k-1}, \text{AGG}_k \left(\{ h_u^{k-1}, \forall u \in N(v) \} \right) \right) \right) \quad (2)$$

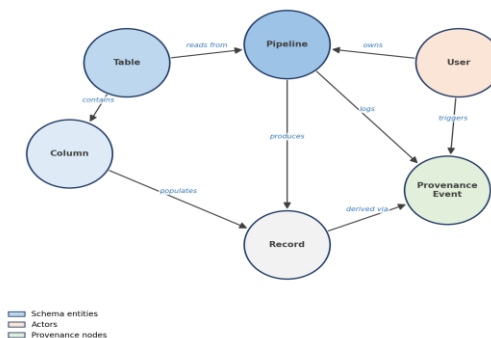


Figure 4

Heterogeneous Enterprise Data Graph Schema, Including Provenance Nodes

Temporal interaction networks are a form of networks in which the entities interact at irregular time points instead of a fixed set of edges, and these networks have been modeled by embedding trajectories that evolve over time as additional interactions are added. One approach suggested in this work uses coupled recurrent update functions to propagate the embeddings of both interacting entities forward in time and predict the future interactions, showing the viability of on-going embedding updates, driven by the events as they happen, when dealing with streaming enterprise data, like pipeline execution logs and access events (Kumar et al., 2019).

4. Dynamic and Temporal Graph Learning for Reliability Assessment

A detailed review of graph anomaly detection techniques based on deep learning has classified the methods based on the type of anomaly to be detected (anomalous nodes, edges, subgraphs, and entire graphs) and on the learning paradigms used (autoencoder-based reconstruction, generative adversarial methods, and contrastive learning). The taxonomy can be directly applied to the classification of unreliable enterprise data artifacts that can be either an anomalous record within an enterprise, an anomalous relationship between two systems, or a systemic degradation of an entire data domain (Ma et al., 2021).

The recent studies on deep graph anomaly detection have highlighted several ongoing challenges: sparse labeling; imbalanced class distribution between normal and abnormal examples; and the fact that anomalous nodes are often defined not by unusual features in isolation but only in the context of their relationships. These observations lead to hybrid architectures that use both structural and attribute-based signals, rather than either alone (Qiao et al.,).

The temporal graph network has been formalized as an encoder-decoder network, where

each node stores a memory vector that is modified by an incident event using a learnable memory function. This update mechanism is represented by Equation 3, where s is the node memory state, t^- is the time just before the current event occurs, and m is a message vector generated from the attributes of the incoming event.

$$s_{i(t)} = \text{mem}(s_{i(t^-)}, m_{i(t)}) \quad (3)$$

This design has been proven to combine a wide variety of existing dynamic graph models and allow efficient training with streamed interaction data, which is well suited to enterprise applications where data reliability needs to be re-evaluated at regular intervals, but not recalculated from scratch each time (Rossi et al., 2020).

Self-attention mechanisms have been generalized to dynamic graphs with architectures that learn structural neighborhood attention within each snapshot as well as temporal self-attention across a sequence of snapshots. The weight of the temporal attention to representations at two time points is computed as in Equation 4, where W_q and W_k represent learned query and key projection matrices and d is the embedding dimensionality used for scaling.

$$\beta_{t,t'} = \text{softmax}_t' \left(\frac{((W_q h_t)^T (W_k h_{t'}))}{\sqrt{d}} \right) \quad (4)$$

This dual-attention design has been found to be able to track the changing structural role better than models using only structural or temporal evidence of unreliability, respectively, indicating that a reliability assessment system should also separate structural and temporal evidence of unreliability, and then recombine them, rather than combining them in a single representation (Sankar et al., 2020). Figure 5 shows the propagation of reliability estimates across entities for each temporal snapshot observed, taking into consideration the attention weights.

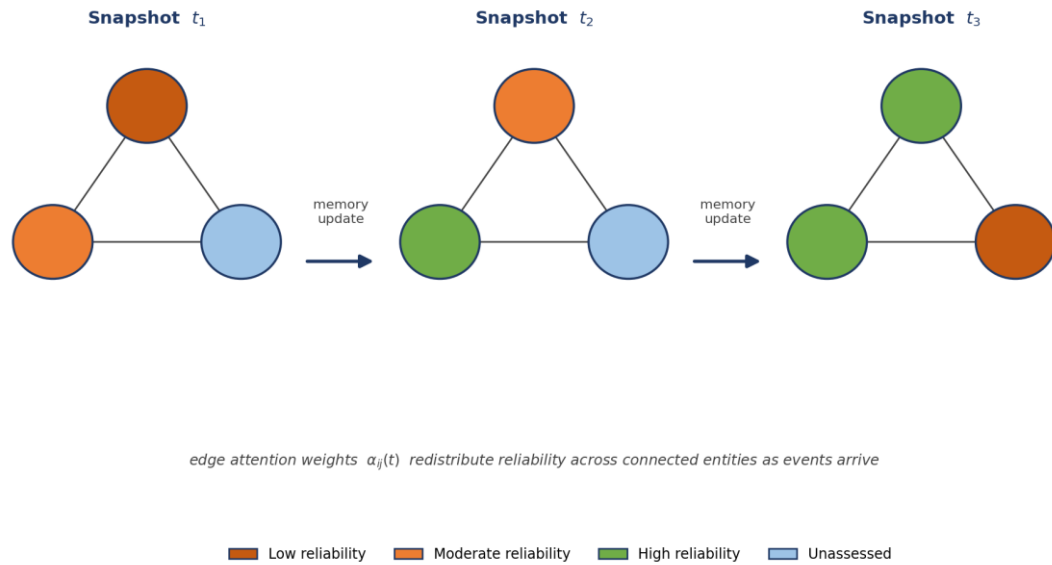


Figure 5

Attention-Weighted Reliability Propagation Across Successive Temporal Graph Snapshots

Table 2

Graph Learning Architectures Reviewed and Their Mechanistic Basis

Reference	Architecture	Core Mechanism	Relevance to Reliability Assessment
Hamilton et al. (2017)	GraphSAGE	Inductive neighborhood sampling and aggregation	Generalizes to newly created enterprise entities
Kipf & Welling (2017)	Graph Convolutional Network	Spectral-approximate normalized adjacency propagation	Efficient semi-supervised reliability inference
Kumar et al. (2019)	JODIE-style embedding	Coupled recurrent trajectory projection	Continuous update from event streams
Rossi et al. (2020)	Temporal Graph Network	Per-node memory updated by incident events	Streaming, low-latency reliability updates
Sankar et al. (2020)	DySAT	Joint structural and temporal self-attention	Separates structural from temporal unreliability signals

5. Proposed DataTrustGraph Framework

The original formulation of the graph neural network model set the stage for the subsequent convolutional and attentional generalizations, by defining a propagation function through contraction that can be iterated until it reaches a fixed point, where the node state is calculated. The DataTrustGraph framework proposed here takes this iterative propagation principle one step further: it uses a metric of reliability, instead of a generic classification label,

as the quantity propagated throughout the enterprise data graph (Scarselli et al., 2009).

Empirical research in secure digital research environments has shown that systematic tracking of data access, transformation and disclosure events aids in the understanding of the auditability and regulatory compliance of such research. The architecture proposed here extends this concept by adding a provenance layer: every transformation event is represented as a typed edge connecting a source data node, a transformation

node, and a resulting data node, and the history of transformations by which each entity has been created can be traversed as well as the relationships between the substantive entities (Scheliga et al., 2022).

An alternative to fixed or degree normalized aggregation schemes has been put forward by attention-based neighborhood aggregation, where an attention coefficient, representing the relative importance of a neighbouring node, is learned instead of being assumed. This coefficient is obtained as shown in Equation 5, where W represents a shared linear transformation, a represents a learnable attention vector, and the concatenation operator refers to the representations of the node being evaluated and each of its neighbors as transformed by W .

$$\alpha_{ij} = \text{softmax}_j(\text{LeakyReLU}(a^T [W h_i \| W h_j])) \quad (5)$$

In the proposed framework, this mechanism is adapted to the fact that the attention coefficients are directly proportional to the influence of a neighboring data entity's reliability in the estimate of the entity whose reliability is being assessed, enabling the model to give less weight to a neighboring data entity that is structurally connected but substantively uninformative about reliability (Veličković et al., 2018).

Knowledge graph embedding methods map entities and relations into a continuous vector space so that relational structure is preserved through geometric or algebraic constraints; translation-based, tensor-factorization-based, and neural methods each carry distinct strengths and limitations (Wang, Mao, et al., 2017). The DataTrustGraph framework is based on these embedding principles for encoding heterogeneous enterprise entities (such as tables, records, transformation pipelines and users) into a single representational space before propagating reliability.

5.1. Architectural Layers

The proposed framework is structured into four cooperating levels: ingestion layer (representing enterprise artifacts as heterogeneous graph nodes); provenance layer (representing a transformation and access event as a typed edge, according to the auditability principles used in secure research environments; Scheliga et al., 2022); representation layer (embedding of nodes using

inductive/attention weighted aggregation); propagation layer (iterative computation of the reliability estimate through the mechanisms described in Sections 4 and 6).

6. Reliability Scoring and Data Quality Dimensions

A conceptual framework of data quality in the data consumers' perspective has been developed that defines four dimensions of data quality: intrinsic, contextual, representational, and accessibility dimensions. In the proposed framework this multidimensional structure is used as the foundation to build node-level intrinsic quality features before propagation in the graph (Wang & Strong, 1996).

Since the graphs can have multiple kinds of nodes and edges, heterogeneous graph attention networks were introduced, which utilize a two-level attention mechanism: node-level attention for nodes in each metapath-based neighborhood, and semantic-level attention for nodes with different metapaths. For the semantic-level combination, each metapath-specific embedding is weighted by a learned semantic-level attention coefficient, and summed across all metapaths under consideration, as shown in Equation 6.

$$Z = \sum_{p=1}^P \beta_p Z_p \quad (6)$$

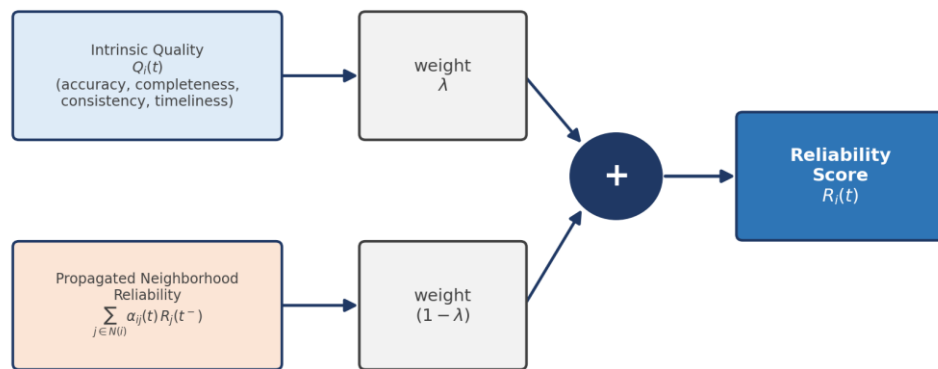
This dual-attention pattern is well-suited to enterprise data graphs, where the graph nodes are typically a diverse collection of entities (tables, columns, pipelines, users) linked by a variety of different types of relations including ownership, transformation, and access relations that can have very different implications for reliability (Wang, Ji, et al., 2019).

A thorough review of the state of the art has been made, providing an organization of graph neural network architectures based on recurrent, convolutional, autoencoder, and spatial-temporal categories, and cataloging common training procedures, benchmark tasks and open research problems. This taxonomy guides the selection of a hybrid spatial-temporal encoder within the proposed framework, an encoder that can be used to encode the static structural backbone and the dynamic event stream that is typical of enterprise data pipelines (Wu et al., 2021). Based on the above intrinsic quality dimensions and attention-weighted propagation mechanisms, a composite reliability score for an entity i at time t is suggested in Equation

7, where Q represents the intrinsic quality score derived from the intrinsic quality dimensions defined in the data quality literature, α represents the dynamic, attention-derived propagation weight associated with the neighbor j , $N(i)$ represents the neighbor set N of entity i , and λ is a tunable

coefficient used to balance between intrinsic quality and propagated neighbor set reliability, with $0 < \lambda < 1$.

$$R_{i(t)} = \lambda \cdot Q_{i(t)} + (1 - \lambda) \cdot \sum_{j \in N(i)} \alpha_{ij(t)} R_{j(t-)} \quad (7)$$



$$\text{Equation 7: } R_i(t) = \lambda \cdot Q_i(t) + (1 - \lambda) \sum_{j \in N(i)} \alpha_{ij(t)} R_j(t-)$$

Figure 6

Conceptual Decomposition of the Composite Reliability Score Defined in Equation 7

This recursive structure of Equation 7 reflects the iterative propagation principle used in the initial formulations of graph neural networks and the explicit intrinsic-quality term ensures direct mapping to previous, human-interpretable notions of data quality dimensions, which is a limitation often highlighted in learned representations. This decomposition is shown schematically in Figure 6, which separates the contributions to the final reliability estimate into an intrinsic component and a structural component.

Table 3
Mapping of Proposed Framework Components to Reviewed Literature

Framework Component	Governing Principle	Primary Reference(s)	Function
Entity and provenance graph	Provenance as first-class graph structure	Kleinsteuber et al., Scheliga et al. (2022)	Captures lineage and transformation history
Representation layer	Inductive, heterogeneous embedding	Hamilton et al. (2017), Wang, Mao, et al. (2017), Wang, Ji, et al. (2019)	Encodes heterogeneous entities into shared vector space
Structural propagation	Attention-weighted neighborhood aggregation	Kipf & Welling (2017), Veličković et al. (2018)	Infers reliability from structurally connected entities
Temporal propagation	Memory- and attention-based event updating	Kumar et al. (2019), Rossi et al. (2020), Sankar et al. (2020)	Maintains continuously updated reliability estimates
Intrinsic scoring	Multidimensional data quality theory	Cichy & Rass (2019), Wang & Strong (1996)	Supplies interpretable, human-auditable quality inputs

7. Discussion: Challenges, Governance, and Future Directions

Scalability, interpretability and robustness to adversarial and noisy inputs have been recurring open challenges throughout reviews of graph neural network methods and applications. Auditable, explainable outputs are essential in enterprise deployments of reliability assessment systems, besides just the scalar score, because they are needed to support trust-related decisions that could have operational or regulatory implications (Zhou et al., 2020).

The deployment is complicated because of governance considerations: continuous reliability scoring requires continued access to operational sensitive data, which poses questions of data minimization, access control, and retention period of provenance logs. The attention-based propagation mechanisms described in Sections 4 through 6 provide a potential avenue for reconciling transparency with confidentiality requirements, but there has been no literature that suggests an empirically tested way to resolve this tension (Zhou et al., 2020).

In the future, incremental graph learning algorithms that allow updating the reliability estimates without retraining the entire graph and enterprise-specific evaluation protocols designed for enterprise data reliability instead of generic anomaly detection evaluation protocols could be added to the proposed framework. These custom protocols would make it easier to compare candidate architectures directly and also create the opportunity to empirically calibrate the composite reliability formulation given in Equation 7 against a definition of acceptable propagation depth and intrinsic quality specific to a particular organization.

8. Conclusion

This paper synthesizes literature on graph anomaly detection, static and dynamic graph representation learning, knowledge graph embedding, data provenance management, and classical data quality theory and proposed a conceptual framework named DataTrustGraph for continuous, graph-based assessment of the reliability of enterprise data. The framework provides a principled alternative to assessing data quality at the record level that is static, by representing enterprise entities, transformation events, and provenance metadata in a unified,

heterogeneous, temporally evolving graph; and by propagating reliability through an attention-weighted mechanism developed from existing graph learning architectures. The composite reliability formulation preserves the interpretability of established data quality dimensions while incorporating dynamically propagated neighborhood evidence, addressing interpretability concerns raised throughout the literature. There is still considerable effort to be put into applying this conceptual synthesis to a validated system, especially when working incrementally in a streaming environment, with standardized evaluation routines for enterprise reliability assessment and with the reconciliation of ongoing provenance monitoring and data governance and access control constraints. Addressing these directions is an essential step towards the practical application of graph-based reliability assessment in enterprise data environments.

References

- [1] Akoglu, L., Tong, H., & Koutra, D. (2015). Graph based anomaly detection and description: A survey. *Data Mining and Knowledge Discovery*, 29(3), 626–688. <https://doi.org/10.1007/s10618-014-0365-y>
- [2] Bellini, E., Iraqi, Y., & Damiani, E. (2020). Blockchain-based distributed trust and reputation management systems: A survey. *IEEE Access*, 8, 21127–21151. <https://doi.org/10.1109/ACCESS.2020.2969820>
- [3] Budach, L., Feuerpfeil, M., Ihde, N., Nathansen, A., Noack, N., Patzlaff, H., Harmouch, H., & Naumann, F. (2022). *The effects of data quality on machine learning performance*. arXiv. <https://doi.org/10.48550/arXiv.2207.14529>
- [4] Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. *Data Science Journal*, 14, Article 2. <https://doi.org/10.5334/dsj-2015-002>
- [5] Cichy, C., & Rass, S. (2019). An overview of data quality frameworks. *IEEE Access*, 7, 24634–24648. <https://doi.org/10.1109/ACCESS.2019.2899751>
- [6] Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Inductive representation learning on

- large graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (pp. 1025–1035). Curran Associates.
- [7] Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *Proceedings of the 5th International Conference on Learning Representations*.
- [8] Kumar, S., Zhang, X., & Leskovec, J. (2019). Predicting dynamic embedding trajectory in temporal interaction networks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 1269–1278). Association for Computing Machinery. <https://doi.org/10.1145/3292500.3330895>
- [9] Ma, X., Wu, J., Xue, S., Yang, J., Zhou, C., Sheng, Q. Z., Xiong, H., & Akoglu, L. (2021). A comprehensive survey on graph anomaly detection with deep learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(12), 12012–12038. <https://doi.org/10.1109/TKDE.2021.3118815>
- [10] Rossi, E., Chamberlain, B., Frasca, F., Eynard, D., Monti, F., & Bronstein, M. (2020). *Temporal graph networks for deep learning on dynamic graphs*. arXiv. <https://doi.org/10.48550/arXiv.2006.10637>
- [11] Sankar, A., Wu, Y., Gou, L., Zhang, W., & Yang, H. (2020). DySAT: Deep neural representation learning on dynamic graphs via self-attention networks. In *Proceedings of the 13th International Conference on Web Search and Data Mining* (pp. 519–527). Association for Computing Machinery. <https://doi.org/10.1145/3336191.3371845>
- [12] Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2009). The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1), 61–80. <https://doi.org/10.1109/TNN.2008.2005605>
- [13] Scheliga, B., Markovic, M., Rowlands, H., Wozniak, A., Wilde, K., & Butler, J. (2022). Data provenance tracking and reporting in a high-security digital research environment. *International Journal of Population Data Science*, 7(3). <https://doi.org/10.23889/ijpds.v7i3.1880>
- [14] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph attention networks. In *Proceedings of the 6th International Conference on Learning Representations*.
- [15] Wang, Q., Mao, Z., Wang, B., & Guo, L. (2017). Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, 29(12), 2724–2743. <https://doi.org/10.1109/TKDE.2017.2754499>
- [16] Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4), 5–33. <https://doi.org/10.1080/07421222.1996.11518099>
- [17] Wang, X., Ji, H., Shi, C., Wang, B., Ye, Y., Cui, P., & Yu, P. S. (2019). Heterogeneous graph attention network. In *The World Wide Web Conference* (pp. 2022–2032). Association for Computing Machinery. <https://doi.org/10.1145/3308558.3313562>
- [18] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2021). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1), 4–24. <https://doi.org/10.1109/TNNLS.2020.2978386>
- [19] Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI Open*, 1, 57–81. <https://doi.org/10.1016/j.aiopen.2021.01.001>