Deriving Uncertain Knowledge in Knowledge Graphs Under Theoretical Mapping of Heuristics

Simon von Oppenkowski, Benedikt Lerch and Klemens Schnattinger
Data Science and Artificial Intelligence, Baden-Wuerttemberg Cooperative State University, Loerrach, Germany

vonoppes@dhbw-loerrach.de
lerchb@dhbw-loerrach.de
schnattinger@dhbw-loerrach.de

Abstract: This paper proposes a deterministic, explainable framework for validating uncertain entities during Knowledge Graph (KG) expansion. The contribution is conceptual and formal: we present the framework and demonstrate its internal mathematical coherence, while empirical evaluation is explicitly deferred to future work. Since adding unverified nodes is trivial but removing structurally corrupted data afterwards is computationally hard, the multiphase Let's Build Knowledge (LBK) Framework acts as a gatekeeper that prevents such corruption at the point of entry. Drawing on cognitive heuristics, molecular data patterns, and meta-reasoning, LBK offers a deterministic alternative to prevalent stochastic and embedding-based approaches in Explainable AI (XAI): it integrates Formal Concept Analysis (FCA) with a localized adaptation of Byzantine Fault Tolerance (BFT) to scrutinize incoming structures, enabling loss-free, rule-based verification of new entities before global integration. We provide a precise specification of this hybrid model and demonstrate its formal capacity for deterministic uncertainty management, without claiming empirical validation.

Keywords: Knowledge, Uncertainty, Subgraphs, Isomorphism, FCA, Hyperedge

1. Introduction

This paper pursues a single, clearly bounded goal: to design and formally justify a method that validates newly added, uncertain entities in a Knowledge Graph without resorting to probabilistic embeddings. In doing so, we extend a quality-based, terminological tradition in which uncertain concept hypotheses are legitimized or refuted by logic-based meta-level assessment rather than stochastic estimation (Hahn and Schnattinger 1998; Schnattinger and Hahn 1996), transferring this principle from text-driven concept acquisition to graph-structured entity validation. Concretely, we (i) propose the Let's Build Knowledge (LBK) Framework, (ii) demonstrate that Formal Concept Analysis and a localized Byzantine consensus can validate uncertain nodes deterministically within a bounded neighborhood, and (iii) specify a transparent, reproducible validation function. In design-science terms, the contributed artifact is a method whose utility is argued analytically; the paper's claim is thus the design and formal justification of the method, not its empirical confirmation (Gregor and Hevner 2013). The remainder of this introduction motivates the need for such a deterministic alternative; Sections 2–3 develop the formal model, whose mathematical coherence is demonstrated while empirical benchmarking remains future work.

The need for such validation arises from a concrete problem. In today's digital transmission of knowledge, inconsistent or false information is increasingly conveyed, leading to severe structural uncertainties (Zhou et al. 2022). Generative AI has accelerated this challenge: AI-generated content amplifies disinformation at scale, making synthetic text and fabricated media hard to distinguish from authentic sources (Cazzamatta and Sarısakaløğlu 2025), regardless of the medium (Broda and Strömbäck 2024). LLMs remain highly prone to hallucination and weak reasoning in domain-specific contexts (Ma et al. 2025), motivating the integration of structured Knowledge Graphs (KGs) for factual grounding and explicit reasoning paths (Zhang, Zhou, and Yang 2025).

KGs are the structural answer to these limitations but introduce a validation problem of their own. Within KGs such as Wikidata (Vrandečić 2012), semantic structures enable critical applications like semantic search (Noy et al. 2019), Graph Neural Networks for recommender systems (Wang et al. 2019), and advanced question answering (Huang et al. 2019; Ma et al. 2025). Yet handling new, unverified data within a semantic triple remains fundamental, as every newly generated node is initially uncertain (Schnattinger and Hahn 1996; Jarnac, Chabot and Couceiro 2025). Current embedding methods address this probabilistically, which limits interpretability, factual precision, and deterministic validation (Cao et al. 2024).

These probabilistic limitations are precisely what the stated goal seeks to avoid; for a deterministic alternative, we turn to cognitive science as a theoretical foundation for uncertainty resolution. Substituting stochastic estimation with a graded, evidence-quality assessment of competing hypotheses has already been operationalized in a purely symbolic, terminological setting (Hahn and Schnattinger 1998), which we take as a direct methodological precedent. In psychology, humans resolve factual uncertainties even without explicit prior knowledge through learned patterns known as eristic reasoning (Kurdoglu, Jekel, and Ateş 2023): such heuristics act as cognitive shortcuts that enable rapid judgment by substituting complex structural assessments with simpler, related properties (Tversky and Kahneman 1974). Because this pattern matching operates independently of formal stochastic reasoning, it is an ideal blueprint for machine-based, explainable uncertainty resolution in knowledge graphs (Spiliopoulos and Hertwig 2024). Inspired by this, we introduce the Let's Build Knowledge (LBK) Framework, which rejects opaque vector embeddings and instead applies meta-heuristic, structural information to deterministically legitimize or refute uncertain knowledge units and question-answer relationships.

Our goal and contributions are guided by one central research question:

Research Question (RQ): How can unverified, newly introduced entities within an expanding Knowledge Graph be validated in a deterministic, explainable, and computationally bounded manner without relying on probabilistic embeddings?

To answer it, the paper makes three theoretical contributions to Explainable AI (XAI), each stated as what we propose, demonstrate, or establish, not as an empirically proven result:

1. Topological Quarantining**:** a method to structurally isolate uncertain data into localized subgraphs, bounding the processing scope to a constant execution radius $2 \leq dist(x, v) \leq 4$.
2. Deterministic Pattern Derivation: showing how Formal Concept Analysis (FCA) and Galois connections map multidimensional node attributes into a lossless formal context lattice.
3. Consensus-Driven Reintegration**:** a mathematical validation function $(\Psi(x))$ based on an algorithmic transfer of Byzantine Fault Tolerance (BFT) to formal concepts, enforcing a strict 75% threshold for deterministic graph expansion.

2. Related Work

2.1 State of the Art

Against the delimitation from LLM-based reasoning, the state of the art cannot be identified unambiguously. Knowledge Graph Completion (KGC) is closely related, as both approaches address incompleteness in knowledge graphs, but their objectives differ fundamentally: KGC estimates the probability of missing relations between already accepted entities and predicts plausible links from learned embeddings (Hou et al. 2025), whereas the proposed approach evaluates whether a previously unknown candidate node should be accepted at all. LBK thus inverts the setting — it does not score the plausibility of a relation but decides the admissibility of a newly arriving, unverified entity before it may participate in the graph: KGC answers which edge is probable, LBK whether the node may enter, a deterministic gatekeeping decision with an auditable justification rather than a probabilistic ranking. A more closely related, logic-based approach is given by Delfino, Lenzerini and Poggi, who define a formal-logical semantics for RDFS-based knowledge graphs (RKGs) — including their metamodeling features — and analyze query answering, showing in particular that such graphs need not possess a universal model (Delfino, Lenzerini, and Poggi 2025).

2.2 Hyper-reasoning

A fundamental problem in abstracting and validating newly added nodes is the purely binary, incident nature of edges: while edges define connections, in isolation they hardly transfer complex information. Ontologies offer a first semantic enrichment by formalizing relationships as RDF triples (Stuckenschmidt 2011). To extend the reasoning chain further, this work breaks up the binary relation between two nodes through hyper-reasoning. Decisively shaped by Gao and Zhang, this technique generates a set of edges together with their attributes and declares them a hyperedge (Gao et al. 2021). Such a hyperedge does not alter the fundamental graph structure but enables a more differentiated perspective: rather than evaluating purely binary constellations, it groups nodes into a common set via connected relations — for the subgraph construction presented here, an essential step to ensure the injective mapping of information.

Even though current analyses partly question the superiority of Hypergraph Neural Networks (HGNN) over conventional Graph Neural Networks (GNN) (Kim et al. 2024), hypergraph structure is fundamental to the reasoning approach chosen here, since without it the correct, injective representation of complex ontological relationships could not be guaranteed. Modern hyper-reasoning performs well given high computational resources but fails in real-time scenarios, as it processes the graph globally — impractical for on-demand decisions. Crucially, our use of hypergraphs differs in kind, not only scale: HGNNs embed higher-order structure into a learned vector space and propagate messages globally (Gao et al. 2021; Kim et al. 2024), drawing their power from this global, statistical aggregation, whereas LBK uses the hyperedge only as a non-learned structural carrier guaranteeing an injective, loss-free grouping of local context, with no embedding or message passing at all. We thus deliberately trade representational expressiveness for locality, determinism, and explainability — a trade-off mainstream hypergraph reasoning does not make. By combining hyper-reasoning with a strict limitation of hops, LBK confines validation to a localized neighborhood $(2 \leq dist(x, v) \leq 4)$, drastically reducing the required computing power and enabling instant, operational decisions.

2.3 Meta-reasoning

Since every newly added node is initially interpreted as uncertain, the definition of "validity" must be derived congruently. To this end, meta-reasoning is applied within the isolated subgraph: because its injective mapping endows every node with a specific set of properties, these can be mapped into property classes via Formal Concept Analysis (FCA) and treated as reasoners on a meta level (Ganter and Wille 2024). A related terminological meta-reasoning perspective is given by Hahn, Klenner and Schnattinger, who formalize meta-level abstractions to validate uncertain or heterogeneous knowledge sources (Hahn, Klenner, and Schnattinger 1996). Methodologically related, Mandal and Medya show that meta-learning in Graph Neural Networks (GNNs) significantly improves speed and accuracy (Mandal et al. 2021) — but where they use statistical embeddings for classification, the present work is purely logic-based. A further line interprets node contents as a meta level to exploit structural (graph-theoretic isomorphism) similarities for reasoning in LLMs, serving a different discipline (Yuan et al. 2025); such works underline the need for meta-level abstractions, which this publication addresses through FCA.

The decisive difference is that the present model is purely symbolic: validation is carried out entirely by FCA and Byzantine consensus, with no learned component anywhere in the pipeline. Neural-symbolic systems gain generalization from their statistical part but inherit its opacity and dependence on training data; LBK relinquishes generalization to keep every validation step inspectable and reproducible, continuing a strictly symbolic line of terminological meta-reasoning (Hahn, Klenner, and Schnattinger 1996). As "meta-reasoning" has no universal definition, FCA is chosen as the abstraction level that formally safeguards validation. LBK also takes a deliberately different stance on uncertainty: whereas surveys describe quantifying and propagating it through probabilities, fuzzy degrees, or confidence weights on triples (Jarnac, Chabot, and Couceiro 2025), LBK resolves it categorically — each candidate is mapped, through local consensus, to a discrete admissibility decision, eliminating uncertainty at the point of entry rather than carrying it forward as a graded annotation. This continues a quality-based terminological tradition in which uncertain hypotheses are graded by discrete quality labels rather than numeric likelihoods (Schnattinger and Hahn 1996).

2.4 Byzantine Fault in Knowledge Graphs

Within this work, a knowledge graph is assumed to represent the mathematical representation of a distributed network. Based on this analogy, the theory of Byzantine Fault Tolerance (BFT) can be transferred to the validation of knowledge structures (Lynch 1997). Here, a network message corresponds to the triple content (semantics) within the ontology, while the edges act as communication paths via which the context (meta information) contributes to consensus formation at the dummy vertex. Byzantine Fault Tolerance is generally represented by the condition $n \geq 3f + 1$; its safety factor of 3 enables a system – in this context the graph – to make an autonomous decision about the validity of information (Pease, Shostak, and Lamport 1980). The original implementation of BFT, however, does not investigate vagueness, and traditional BFT approaches rely strictly on Boolean indicators to verify nodes, making them incapable of evaluating complex, non-binary semantic structures. Due to the increased vagueness within heuristic approaches, a safety factor of 4 is chosen in the further course of this work $(n \geq 4f + 1)$. This deliberate expansion of the fault-tolerance ceiling counters the non-binary entropy of semantic data, providing the formal mathematical justification for the consensus thresholds executed in the subsequent validation phases.

2.5 Provenance and Delimitation of the Dummy-vertex Model

The use of so-called dummy nodes originates primarily in medical research, mainly within computer-aided antibiotic modelling, where the dummy node represents a potential new molecule within an existing chemical application chain. Mathematically, an initially empty node is added to an original graph G, its interaction potential defined by the topological neighborhood and the corresponding edge relations $E \subseteq V \times V$. Since the empty node initially exerts no intrinsic influence on the overall system, it allows the isolated simulation of connections to target nodes.

Complementarily, the dummy- or virtual-node construct itself has a direct precedent in graph-based machine learning. Pham et al. augment an attributed graph with an initially empty node connected to the existing nodes, whose representation captures latent aspects not immediately available from the local attributes and connectivity structure (Pham et al. 2017). Because this construct was validated, among other tasks, on predicting the bioactivity of chemical compounds, it substantiates the isolation-and-trial-embedding principle that the present work projects onto knowledge-graph validation. A closely related line in biochemistry applies such trial embeddings to assess the influence of altered active substances on a target node (Gilmer et al. 2017). Methodologically, however, the present approach differs clearly: rather than the learned message passing and vectorization used there to derive Euclidean embeddings, the LBK Framework employs the dummy node solely as a non-learned structural carrier for local, logic-based legitimation.

Further differentiation concerns Cui and Abdelzaher (2024), where dummy nodes serve as anchor points to approximate incomplete knowledge via virtual paths; in contrast, the method presented here does not focus on statistical approximation but on the explicit validation of vague knowledge units through Byzantine consensus.

3. Formal Representation

3.1 Formal-theoretic Approach

Assumptions. Before the formalization, we fix the assumptions used throughout. (A1) Every newly arriving entity x is initially treated as uncertain and must be validated before global integration. (A2) The accepted graph $G = (V, E)$ is regarded as a trusted distributed system, so that a Byzantine-style consensus over neighbouring evidence is meaningful. (A3) Information transfer from G into the local subgraph is injective and therefore loss-free. (A4) Only the bounded neighbourhood $2 \leq dist(x, v) \leq 4$ is consulted; a candidate with no admissible neighbor is rejected. (A5) Attributes are discrete, so that the incidence relation of the formal context is well-defined and the Galois derivations are exact. These assumptions delimit the scope of every definition and result below.

This work validates uncertainties in a knowledge graph by transferring heuristic ways of thinking to a machine. This requires both establishing the factorization for validation and validating the node resource-efficiently even within a large graph. To this end, a new node is inserted into an existing graph, an attribute signature is derived in its context, and — via this signature and a fixed number of neighbors — a validation set of nodes is spanned. To map this set context specifically, the hyperedge approach is used; within it, the edge contents follow transitivity and reflexivity derived from the established attribute signature. The spanned space is limited to a maximum neighborhood of 4 nodes, while a neighborhood must nonetheless be present, since a disjoint node implicitly corresponds to untruth and cannot be incorporated into the original graph.

This validation space of edge and node relationships is mapped within a subgraph, guaranteeing performance and allowing large amounts of data to be considered in an isolated, context-specific manner; due to the injective property of the graph, information transfers into the subgraph without loss. There, Formal Concept Analysis serves as the validation operator: by the Galois derivation applied to the semantic sets according to Ganter and Wille, the common attributes between the surrounding nodes and the node to be validated are derived via intersection. For robustness Byzantine Fault Tolerance then validates this intersection against the totality of attributes, which must reach a threshold of at least 75%; meeting it validates the new node within its context and renders it robust with respect to the graph.

If the dummy node is classified as robust, it can be integrated into the original graph. The reason for this is that subgraphs can always be traced back to the original graph.

3.2 Mathematical Representation

Table of Annotation		
Annotation	Definition	Context
$\boldsymbol{G = (V, E)}$	Representation of KG	Existing as true accepted KG
$\boldsymbol{x = V_{dum}}$	Vague Dummy Node	New to enter and unverified entity

Table of Annotation		
$\boldsymbol{X}$	Isolated Subgraph set	Set of neighborhood hops within the radius $2 \leq dist(x, v) \leq 4$
$\boldsymbol{K = (O, A, I)}$	Formal context (FCA)	O = Object, A = Attributes, Incidence relation $I \subseteq O \times A$
$\boldsymbol{M_x}$	Attribute set of x	Set of Attributes contained in x
$\boldsymbol{M'_x}$	First Galois derivation	Set of objects O that consistently contain all attributes of M_x
$\boldsymbol{M''_x}$	Second Galois derivation	Set of attributes A consistently shared by the neighborhood objects in M'_x
$\boldsymbol{n}$	Cardinality of M''_x	Total cardinality of the validated attribute consensus set $n = \lvert M''_x \rvert$
$\boldsymbol{f_{obs}}$	Attribute deficit	Number of missing or deviating attributes $\lvert M''_x \setminus M_x \rvert = n - \lvert M_x \rvert$
$\boldsymbol{\Psi}(\boldsymbol{x})$	Validation function	Calculated consensus score for the dummy node
$\boldsymbol{\Gamma}$	Absolute threshold	Baseline barrier for reintegration (initially set to 0.75)

The fundamental formal representation of an original graph is henceforth defined as $G = (V, E)$. It should be noted beforehand that the conventional graph structure as well as the graph modification refer to the teachings of Reinhard Diestel and all approaches are taken from the book Graph Theory (Diestel 2025). As usual, $e \in E$ represents an edge, while $v \in V$ represents a node in a graph. The dummy vertex is expressed in a simplified manner as x, or $V_{dum} = x$. To integrate x into an existing graph, the following holds:

$$G' = (V', E') \mid V' = V \cup \{x\}, E' = E \cup \{e_x\}$$

In general, the neighborhood relation is defined over the edge set e_x, where $e_x = e_{x_N}$. The attribute signature of the candidate node $x$is denoted by $\Sigma_{\mathrm{Features_{dummy}}} = M_x$. Following Diestel, neighborhoods can be characterized by shortest-path distances. In this work, however, the direct neighborhood is defined as

$$N(x) := \{v \in V \mid \mathrm{label}(x, v) = \mathrm{true}\}.$$

Furthermore, the admissible hull of x is defined as

$$H(x) := \{v \in V \mid 2 \leq \mathrm{dist}(x, v) \leq 4\}.$$

The direct 1-hop neighborhood is intentionally excluded from the validation process to avoid trivial confirmation and circular validation effects. Consequently, only vertices at distances two to four contribute to the derivation of the consensus attribute set. Since the hull of the subgraph and the relationship of the nodes within this subset have now been explained, the mapping m is established in the following. Diestel describes the mapping of an injective graph by $G' \subseteq G$. Since, in the present case, G' is declared through the addition of the information, the mapping follows as

Let: $m: G' \to G''$ whereby m is strictly bound to the neighborhood filter $H(x)$.

Substructure mapping (clarification). To avoid overloading the symbol G, we write $G'' \preccurlyeq G' \subseteq G$: G' is the graph after tentatively inserting x, and G'' the local validation subgraph. Inside G'' every node v carries its attribute set $M_v \subseteq A$ and every edge its incidence, and I is restricted to $V(G'') \times A$; this restriction is what makes the Galois machinery applicable locally.

The construction of the subgraph G'' is at the same time the activation function of the validation process; therefore G'' is equated. Within the local subgraph $G'' = (V'', E'')$, the formal context $K'' = (O, A, I)$ is induced by setting

$$O := V'', \qquad A := \bigcup_{v \in V''} M_v, \qquad I := \{(v, a) \in O \times A \mid a \in M_v\}.$$

This construction makes the incidence relation well-defined and the Galois derivations on K'' exact, so that the application of Galois theory to the nodes and edges within the subgraph is ensured.

We assume $x \notin O$. If x were already contained in the formal context the validation function would evaluate trivially to $\Psi(x) = 1$.

Now, the node to be validated is defined as x with the features M_x, and the validation set as M_x''. This step leads to the following result: M_x is the set of attributes of x. Through the first Galois derivation M_x, the set of matching nodes from the total set G in the subgraph G'' is formed. Subsequently, M_x is checked against M_x', whereby the result would have to equal 1 for an equality. Due to ontological vagueness, reaching 1 is nearly impossible, which is why ≈ 1 represents a perfect result. For the case that a perfect result is not possible, the approximation and the threshold are defined in the following:

$$M' = \{o \in O \mid \forall a \in A(x) : \{o, a\} \in I\}$$

$$M'' = \{a \in A \mid \forall o \in M' : \{a, o\} \in I\}$$

The validation function is defined as

$$\Psi: \{x\} \to [0,1],$$
$$\Psi(x) := \frac{|M_x \cap M_x''|}{|M_x''|}.$$

If $M_x{}'' = \emptyset$, the validation is undefined and therefore $\Psi(x) := 0$.

The admissible deviation is derived from the Byzantine fault tolerance assumption

$$n \geq 4f + 1,$$

where

$$n := |M_x''|$$

denotes the size of the consensus attribute set and f the maximum tolerated deviation. Solving for f yields

$$f_{max} \leq \frac{n-1}{4}.$$

The corresponding relative deviation threshold is defined as

$$\Gamma' := \frac{f_{max}}{n}.$$

Consequently, a candidate node is automatically accepted if

$$\Psi(x) \geq 1 - \Gamma' \;\; or \;\; \Psi(x) \geq \Gamma.$$

where Γ' accounts for discrete boundary effects in small attribute sets.

For sufficiently large values of n, this corresponds to the practical threshold

$$\Psi(x) \geq \Gamma.$$

If the condition is satisfied, the candidate node is integrated into the graph:

$$G_{\text{new}} = (V \cup \{x\}, E \cup \{e_x\}).$$

Otherwise,

$$G_{\text{new}} = G.$$

3.3 Adaptive Decision Policy

One of the existing problems is still the error handling of the threshold value. To be able to address such an issue in the given problem setting, the proposal of adaptive logic is implemented. A candidate is admitted automatically (unsupervised) if $\Psi(x) \geq 0.75$, routed to human expert review (supervised) if $0.5 < \Psi(x) < 0.75$, and rejected (isolation) otherwise, i.e. if $\Psi(x) \leq 0.5$. Formally:

$$\begin{aligned} \Psi(x) \in [0.75,\ 1] \quad &\Rightarrow \text{ unsupervised}, \\ \Psi(x) \in (0.5,\ 0.75) \quad &\Rightarrow \text{ supervised}, \\ \Psi(x) \in [0,\ 0.5] \quad &\Rightarrow \text{ isolation}. \end{aligned}$$

4. Conclusion and Critical Reflection

In summary, this section presented a deterministic, semantics-based framework for relationship validation in Knowledge Graphs. By combining Formal Concept Analysis (FCA) with localized Byzantine consensus, the LBK model derives ontological structure that preserves local context without opaque probabilistic embeddings, and by confining evaluation to a tightly bounded neighborhood it keeps runtimes low and scales to high-dimensional graphs across domains without domain-specific retraining.

A critical reflection nonetheless exposes two structural dependencies. First, capping the search radius at four hops makes the method vulnerable to data sparsity: in fragmented regions the local subgraph may not yield enough objects ($|G|$) and attributes ($|M|$) for Galois derivation to expose meaningful symmetries; future work will therefore explore adaptive radius scaling. Second, the strict absolute threshold (0.75) guards against semantic vagueness but, as a static barrier in non-binary semantic spaces, risks rejecting valid yet weakly connected knowledge units, which motivates investigating symmetric similarity matching between node literals to accommodate gray areas of vagueness.

To test these boundaries empirically, upcoming work will use the GDELT benchmark, injecting flagged true and false information into the local topology to measure runtime, structural stability, and validation precision, and will extend the multiphase architecture with a formalized Human-in-the-Loop verification step to bridge strict algorithmic consensus and nuanced contextual semantics.

Complexity. Because validation is confined to the bounded neighborhood $2 \leq dist(x, v) \leq 4$, the formal context has at most N_{max} objects and $|A|$ attributes, both independent of the global graph size $|V|$. Computing the two Galois derivations is $O(N_{max} \cdot |A|)$, and materializing the local concept structure follows the known bounds for concept computation, which depend on the density of the context rather than on $|V|$ (Kuznetsov and Obiedkov 2002). The per-insertion cost is therefore constant with respect to the global graph; the method scales to large KGs by construction, and the cost is paid locally.

Scalability and worst case. The constant-per-insertion cost assumes sparse-to-moderate neighborhoods. In the worst case – a densely connected hub with a large admissible neighborhood and a dense context – the number of formal concepts can grow exponentially in $|A|$ (Kuznetsov and Obiedkov 2002), so validations near hubs may become expensive even though global scaling is preserved. Bounding the radius mitigates but does not remove this; an attribute cap or a concept-count budget would be required for hard real-time guarantees.

Failure cases. Beyond sparsity, three failure modes deserve attention. (i) Generic-attribute admission: a candidate declaring only a few widely shared attributes can attain $\Psi(x) = 1$ trivially, because its small attribute set is already closed; the consensus then certifies under-specification rather than correctness. (ii) Adversarial attribute stuffing: a malicious node can copy the attributes prevalent in a target neighbourhood to manufacture consensus, since the test rewards agreement, not provenance. (iii) Homogeneous neighbourhoods: where neighbours are near-identical, almost any candidate passes and the discriminative power of Ψ collapses. These cases show that the consensus test certifies local structural conformity, not factual truth – an intrinsic limitation of any purely structural validator.

Limitation of the BFT analogy**.** The Byzantine Fault Tolerance analogy is a theoretical modelling assumption rather than an absolute guarantee. Since related attributes in a Knowledge Graph tend to co-occur, the 75% threshold serves as a principled heuristic. The formal study of consensus under such correlated evidence is therefore left to future work.

AI Declaration: An AI tool (Large Language Model) was used during the development of this paper specifically to assist with the structural formatting and stylistic alignment of the document according to the conference requirements. The AI tool was utilized to reformat existing text blocks and mathematical formalizations into the required Word structure. All conceptual ideas, logical derivations, and the final intellectual content were produced and verified by the author.

Ethics Declaration: The author(s) confirm that ethical clearance was not required for the research presented in this paper, as the study involved the theoretical development of a mathematical framework and did not involve human participants, animal subjects, or the collection of sensitive personal data. All research was conducted in accordance with standard academic integrity guidelines.

References

Broda, E. and Strömbäck, J. (2024) "Misinformation, Disinformation, and Fake News: Lessons from an Interdisciplinary, Systematic Literature Review", *Annals of the International Communication Association*, Vol. 48, No. 2, pp 139–166.

Cao, J., Fang, J., Meng, Z. and Liang, S. (2024) "Knowledge Graph Embedding: A Survey from the Perspective of Representation Spaces", *ACM Computing Surveys*, Vol. 56, No. 6, pp 1–42.

Cazzamatta, R. and Sarısakaløğlu, A. (2025) "AI-Generated Misinformation: A Case Study on Emerging Trends in Fact-Checking Practices Across Brazil, Germany, and the United Kingdom", *Emerging Media*, Advance online publication, https://doi.org/10.1177/27523543251344971.

Cui, H. and Abdelzaher, T. (2024) “Virtual Node Generation for Node Classification in Sparsely-Labeled Graphs”, *Proceedings of the IEEE International Conference on Big Data*, IEEE Computer Society, Washington, DC, pp 1142–1151.

Delfino, R.M., Lenzerini, M. and Poggi, A. (2025) “RDFS Knowledge Graphs Through the Lens of Logic”, *Proceedings of the 28th European Conference on Artificial Intelligence (ECAI)*, IOS Press, Amsterdam, pp 1511–1518.

Diestel, R. (2025) *Graph Theory*, Springer, Berlin, Heidelberg.

Ganter, B. and Wille, R. (2024) *Formal Concept Analysis: Mathematical Foundations*, Springer Nature, Cham.

Gao, Y., Zhang, Z., Lin, H. et al. (2021) “Hypergraph Learning: Methods and Practices”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 44, No. 5, pp 2548–2564.

Gilmer, J., Schoenholz, S. Riley, P. et al. (2017). “Neural Message Passing for Quantum Chemistry.” *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 70:1263–1272.

Gregor, S. and Hevner, A.R. (2013) “Positioning and Presenting Design Science Research for Maximum Impact”, *MIS Quarterly*, Vol. 37, No. 2, pp 337–355.

Hahn, U., Klenner, M. and Schnattinger, K. (1996) “Automated Knowledge Acquisition Meets Metareasoning: Incremental Quality Assessment of Concept Hypotheses during Text Understanding”, *Proceedings of the 10th Knowledge Acquisition for Knowledge-Based Systems Workshop (KAW’96)*, Banff, Canada, July, pp 9–14.

Hahn, U. and Schnattinger, K. (1998) “Towards Text Knowledge Engineering”, *Proceedings of the 15th National Conference on Artificial Intelligence (AAAI '98)*, AAAI Press / The MIT Press, Madison, WI, pp 524–531.

Hou, Y., Zhou, S., Liang, K. et al. (2025) “Soft Reasoning Paths for Knowledge Graph Completion”, *Proceedings of the 34th International Joint Conference on Artificial Intelligence (IJCAI-25)*, Montreal, Canada, July, pp 2937–2945.

Huang, X., Zhang, J., Li, D. et al. (2019) “Knowledge Graph Embedding Based Question Answering”, *Proceedings of the 12th ACM International Conference on Web Search and Data Mining (WSDM)*, Association for Computing Machinery, New York, NY, pp 105–113.

Jarnac, L., Chabot, Y. and Couceiro, M. (2025) “Uncertainty Management in the Construction of Knowledge Graphs: A Survey”, *Transactions on Graph Data and Knowledge*, Vol. 3, No. 1, pp 3:1–3:48.

Kim, S., Lee, S.Y., Gao, Y. et al. (2024) “A Survey on Hypergraph Neural Networks: An In-Depth and Step-By-Step Guide”, *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, New York, NY, pp 6534–6544.

Kurdoglu, R.S., Jekel, M. and Ateş, N.Y. (2023) “Eristic Reasoning: Adaptation to Extreme Uncertainty”, *Frontiers in Psychology*, Vol. 14, Article 1004031, https://doi.org/10.3389/fpsyg.2023.1004031.

Kuznetsov, S. and Obiedkov, S. (2002) “Comparing Performance of Algorithms for Generating Concept Lattices”, *Journal of Experimental & Theoretical Artificial Intelligence*, Vol. 14, No. 2-3, pp 189–216.

Lynch, N.A. (1997) *Distributed Algorithms*, Morgan Kaufmann, San Francisco, CA.

Mandal, D., Medya, S., Uzzi, B. and Aggarwal, C. (2021) “Meta-Learning with Graph Neural Networks: Methods and Applications”, online, arXiv preprint, arXiv:2103.00137, https://arxiv.org/abs/2103.00137.

Noy, N., Gao, Y., Jain, A. et al. (2019) “Industry-Scale Knowledge Graphs: Lessons and Challenges”, *Communications of the ACM*, Vol. 62, No. 8, pp 116–124.

Pease, M., Shostak, R. and Lamport, L. (1980) “Reaching Agreement in the Presence of Faults”, *Journal of the ACM*, Vol. 27, No. 2, pp 228–234.

Pham, T., Tran, T., Dam, H. and Venkatesh, S. (2017) "Graph Classification via Deep Learning with Virtual Nodes", online, arXiv preprint, arXiv:1708.04357, https://arxiv.org/abs/1708.04357.

Schnattinger, K. and Hahn, U. (1996) “A Terminological Qualification Calculus for Preferential Reasoning Under Uncertainty”, in Görz, G. and Hölldobler, S. (eds.) KI-96: Advances in Artificial Intelligence, Springer, Berlin, Heidelberg, pp 349–362.

Spiliopoulos, L. and Hertwig, R. (2024) “Stochastic Heuristics for Decisions under Risk and Uncertainty”, *Frontiers in Psychology*, Vol. 15, Article 1438581, https://doi.org/10.3389/fpsyg.2024.1438581.

Stuckenschmidt, H. (2011) *Ontologien: Konzepte, Technologien und Anwendungen*, Springer, Berlin, Heidelberg.

Tversky, A. and Kahneman, D. (1974) “Judgment under Uncertainty: Heuristics and Biases”, *Science*, Vol. 185, No. 4157, pp 1124–1131.

Vrandečić, D. (2012) “Wikidata: A New Platform for Collaborative Data Collection”, *Proceedings of the 21st International Conference on World Wide Web (WWW)*, Association for Computing Machinery, New York, NY, pp 1063–1064.

Wang, X., He, X., Wang, M. et al. (2019) “Neural Graph Collaborative Filtering”, *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, Association for Computing Machinery, New York, NY, pp 165–174.

Yuan, X., Di, S., Tang, J. et al. (2025) “Towards Self-Cognitive Exploration: Metacognitive Knowledge Graph Retrieval Augmented Generation”, online, arXiv preprint, arXiv:2508.09460, https://arxiv.org/abs/2508.09460.

Zhang, H., Zhou, L. and Yang, H. (2025) “Learning to Retrieve and Reason on Knowledge Graph through Active Self-Reflection”, online, arXiv preprint, arXiv:2502.14932, https://arxiv.org/abs/2502.14932.

Zhou, X., Shu, K., Phoha, V.V. et al. (2022) “This Is Fake! Shared It by Mistake: Assessing the Intent of Fake News Spreaders”, *Proceedings of the ACM Web Conference 2022 (WWW '22)*, Association for Computing Machinery, New York, NY, pp 3685–3694.