GeoKG-Enabled Similarity Computation with Defeasible Reasoning

Bongjae Kwon[1] and Kiyun Yu[2, *]

[1]Department of Civil and Environmental Engineering, Seoul National University, Republic of Korea
[2]Institute of Engineering Research, Seoul National University, Republic of Korea
[*]Corresponding author: kiyun@snu.ac.kr

April 10, 2025

Abstract

This paper investigates the application of defeasible reasoning within geospatial knowledge graphs (GeoKGs) for geospatial similarity computation. Motivated by the need for accurate and interpretable similarity assessments in domains such as urban planning and location-based services, this study proposes a novel approach that combines the structured data representation of GeoKGs with the uncertainty-aware inference capabilities of defeasible logic. A GeoKG is constructed by integrating data from OSMnx, Wikipedia, and GeoNames. Defeasible rules are generated to capture contextual and functional similarities, and a reasoning engine infers similarity scores through priority-based conflict resolution. The proposed method is benchmarked against knowledge graph embedding (KGE) models and a large foundation model (Gemini) using an expert-annotated dataset. While the KGE model achieved 72.3% accuracy and the LFM 68.1%, defeasible reasoning achieved 67.2%. Despite its lower accuracy, it offers superior interpretability by explicitly representing the rationale behind similarity assessments. This transparency is critical in decision-making scenarios where trust and justification are paramount. The study also highlights the impact of rule refinement and conflict resolution strategies on performance, suggesting potential for further improvement. By introducing defeasible reasoning into GeoKG-based similarity computation, this work provides a promising, explainable alternative to black-box models, paving the way for future hybrid approaches that balance accuracy and interpretability.

Keywords

Geospatial Similarity Computation, Defeasible Reasoning, Geospatial Knowledge Graph, Explainable AI, Conflict Resolution

Acknowledgment

This research was conducted with the support of the Ministry of Land, Infrastructure, and Transport (MOLIT) / Korea Agency for Infrastructure Technology Advancement (KAIA) (Project No. RS-2022-00143336, NTIS Project No. 1615013084).

1 Introduction

The accurate and efficient determination of similarity between geospatial entities is a fundamental challenge with broad implications across numerous domains [**1**]. From urban planning and resource management to location-based services and geographic data integration, the ability to quantify the relatedness of spatial objects is paramount. Consider, for instance, the task of identifying functionally similar commercial districts within a city to inform strategic investment decisions. Or, envision a location-based recommendation system suggesting points of interest based on a user's past preferences and the characteristics of nearby locations. In each of these scenarios, the quality and utility of the outcome hinge directly on the accuracy with which we can assess geospatial similarity [**2**]. However, the inherent complexity of geographic space and the multifaceted nature of spatial entities present significant obstacles to achieving this goal [**3**]. Traditional approaches often rely on simple geometric measures such as Euclidean distance or topological relationships, which fail to capture the rich semantic and contextual information that shapes our understanding of spatial similarity [**4**]. Moreover, the increasing availability of heterogeneous geospatial data sources, while promising, introduces challenges related to data integration, consistency, and handling uncertainty [**5**].

The task of geospatial similarity computation is inherently difficult due to several factors. First, geospatial entities are complex objects characterized by a multitude of attributes, relationships, and spatial properties [**6**]. A building, for example, can be described by its geometry, function, architectural style, historical significance, and proximity to other entities [**7**]. Capturing this multifaceted nature requires integrating information from diverse sources and representing it in a coherent and meaningful way. Second, spatial relationships are often context-dependent and influenced by implicit knowledge [**8**]. The similarity between two parks, for instance, might depend on their size, accessibility, the presence of specific amenities (e.g., playgrounds, walking trails), and the surrounding neighborhood context. Formalizing these contextual dependencies and incorporating them into similarity models is a non-trivial task. Third, geospatial data is often incomplete, inconsistent, or uncertain. Data sources may have varying levels of accuracy, coverage, and timeliness, leading to conflicting information about the same entity [**9**]. Dealing with these data quality issues and reasoning under uncertainty is crucial for robust similarity computation. Finally, the sheer scale of geospatial datasets poses computational challenges [**10**]. Efficiently processing and comparing millions of spatial entities requires scalable algorithms and data structures.

To address these challenges, we propose a novel approach that combines the structured knowledge representation of GeoKGs with the flexible inference capabilities of defeasible logic [**11**]. GeoKGs provide a powerful framework for integrating heterogeneous geospatial data, representing entities and their relationships in a semantic and machine-readable format. By leveraging ontologies and knowledge representation techniques, GeoKGs enable us to capture the multifaceted nature of geospatial entities and their contextual dependencies [**12**]. Defeasible logic, on the other hand, provides a non-monotonic reasoning framework that allows us to reason with incomplete and conflicting information [**13**]. By representing similarity relationships as defeasible rules, we can capture the uncertainty and context-dependence inherent in spatial similarity judgments [**14**]. Our approach consists of several key steps:

(1) constructing a GeoKG by integrating data from OSMnx, Wikipedia, and GeoNames
(2) generating a set of defeasible rules to capture contextual and functional similarities between entities
(3) applying these rules using a defeasible reasoning engine to infer similarity scores
(4) resolving potential conflicts between rules through priority-based strategies

The core contribution of this paper lies in the synergistic combination of GeoKGs and defeasible reasoning for geospatial similarity computation. We demonstrate how the structured knowledge representation of GeoKGs can be effectively combined with the flexible inference capabilities of defeasible logic to address the challenges of incomplete information, contextual dependencies, and conflicting evidence. By explicitly representing similarity relationships as defeasible rules, we provide a more interpretable and explainable alternative to black-box machine learning models [**15**]. Furthermore, our approach offers a principled way to resolve conflicts between different similarity criteria, allowing us to capture the nuanced and context-sensitive nature of spatial similarity judgments [**16**]. We applied an existing method for generating defeasible rules in a new way tailored to the geospatial domain, based on spatial proximity, shared attributes, and domain knowledge [**17**]. These rules are designed to capture commonsense reasoning about similarity. For example, a rule might state that "USUALLY, if two restaurants are located near each other and have similar cuisine types, then they are similar." The "USUALLY" quantifier indicates that the rule is defeasible, meaning that it can be overridden by conflicting evidence. We also implement and evaluate different conflict resolution strategies, such as prioritizing rules based on their specificity or credibility.

To verify the effectiveness of our approach, we conducted a series of experiments using real-world geospatial data from Amsterdam. We compared the accuracy of our defeasible reasoning approach against knowledge graph embedding models and LFM-based models, using a ground truth dataset created with expert annotations. The experimental results showed that while the knowledge graph embedding models achieved higher accuracy overall, our defeasible reasoning approach provided a more interpretable and explainable alternative, achieving competitive performance. Specifically, the knowledge graph embedding model achieved an accuracy of approximately 72.3%, while the defeasible reasoning model achieved an accuracy of 67.2%, and the LFM model achieved an accuracy of 68.1%. These results highlight the trade-off between accuracy and interpretability in geospatial similarity computation. Furthermore, we analyzed the impact of different rule refinement and conflict resolution strategies on the overall accuracy and coherence of our approach. Our findings provide insights into the design and implementation of effective defeasible reasoning systems for GeoKG-based applications.

In conclusion, the primary contributions of this paper can be summarized as follows:

- We introduce a novel approach for geospatial similarity computation that combines the structured knowledge representation of GeoKGs with the flexible inference capabilities of defeasible logic.
- We present a method for generating defeasible rules based on spatial proximity, shared attributes, and domain knowledge.
- We implement and evaluate different conflict resolution strategies for resolving conflicts between defeasible rules.
- We conduct a series of experiments using real-world geospatial data to compare the accuracy of our approach against knowledge graph embedding models and LFM-based models.
- We provide a detailed analysis of the trade-offs between accuracy and interpretability in geospatial similarity computation.

In terms of future work, several promising avenues for research remain. First, we plan to investigate more sophisticated rule refinement techniques to improve the accuracy and coverage of our defeasible rule set [**18**, **19**]. This includes exploring methods for automatically learning rules from data and incorporating feedback from domain experts [**20**]. Second, we aim to develop more advanced conflict resolution strategies that take into account the context and credibility of different information sources [**21**]. This could involve using argumentation-based reasoning techniques to resolve conflicts in a more nuanced and transparent way. Third, we intend to explore the use of our approach in other geospatial applications, such as location-based recommendation systems and geographic data integration. Finally, we plan to investigate the scalability of our approach to larger datasets and more complex GeoKGs [**22**]. By addressing these challenges, we can further enhance the effectiveness and applicability of defeasible reasoning for geospatial similarity computation. We believe that our work provides a solid foundation for future research in this area and opens up new possibilities for leveraging the power of GeoKGs and defeasible logic to solve real-world geospatial problems. For example, the use of semantic trajectory data coupled with defeasible reasoning could allow for better modeling of human mobility patterns and improved location-based services. Or consider the use of our approach for emergency response, where rapid and accurate similarity assessment of affected areas is crucial for efficient resource allocation and disaster management.

2 Background

This section provides the necessary background for understanding the concepts and techniques used in this paper. We will cover the fundamentals of GeoKGs, defeasible reasoning, and geospatial similarity computation. We will also introduce the problem setting and notation used throughout the paper.

Knowledge Graphs A Knowledge Graph (KG) is a structured representation of knowledge consisting of entities, concepts, and relationships between them [**23**]. Formally, we defined a KG as a tuple $G = (E, R, F)$, where E is a set of entities, R is a set of relations, and $F \subseteq E \times R \times E$ is a set of facts [**24**, **25**]. Each fact $(e_1, r, e_2) \in F$ represents a relationship r between entities e_1 and e_2. Knowledge Graphs provide a powerful framework for representing and reasoning about complex domains, and have been successfully applied in various applications, including search engines, question answering, and recommendation systems. In the context of geospatial data, a GeoKG is a KG that specifically focuses on representing geospatial entities and their relationships. These entities can include locations, buildings, landmarks, and other geographic features. The relationships between these entities can include spatial relationships (e.g., "nearby", "contains"), functional relationships (e.g., "serves", "locatedIn"), and semantic relationships (e.g., "isA", "relatedTo"). GeoKGs are often built by integrating data from various sources, such as OpenStreetMap (OSM), Wikipedia, and GeoNames [**26**]. This integration process in-

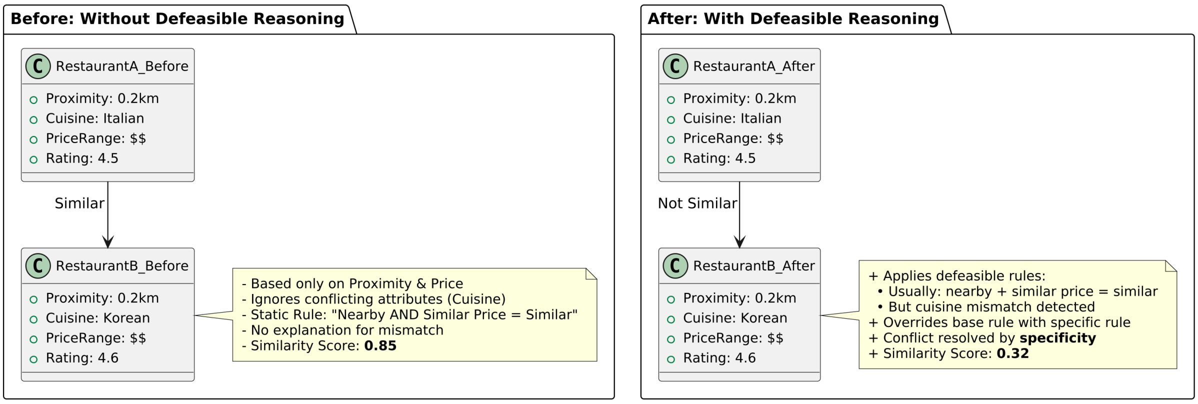

Figure 1: Comparison of Static and Defeasible Reasoning Systems

volves aligning entities across different sources, resolving conflicts, and enriching the KG with additional information.

Defeasible Reasoning Defeasible reasoning is a non-monotonic reasoning approach that allows for reasoning with incomplete and conflicting information [**11**, **27**]. It is based on the concept of defeasible rules, which are rules that can be defeated by contrary evidence. Formally, a defeasible rule can be represented as $A \Rightarrow C$, where A is the antecedent (or body) of the rule, and C is the consequent (or head) of the rule. The "$\Rightarrow$" symbol indicates that the rule is defeasible, meaning that it can be overridden by other rules with stronger evidence. In addition to defeasible rules, defeasible logic also includes strict rules, which are rules that cannot be defeated. A strict rule can be represented as $A \rightarrow C$, where the "$\rightarrow$" symbol indicates that the rule is strict. Defeasible reasoning systems typically include a set of inference rules that specify how defeasible and strict rules can be used to derive conclusions [**28**]. These inference rules often involve mechanisms for resolving conflicts between rules, such as prioritizing rules based on their specificity or credibility. For example, consider the following defeasible rules:

$$\text{bird}(X) \Rightarrow \text{flies}(X)$$

$$\text{penguin}(X) \Rightarrow \neg\text{flies}(X)$$

These rules state that birds usually fly, but penguins usually do not fly. If we know that Tweety is a bird and a penguin, then these rules conflict. To resolve this conflict, we can introduce a priority relation that states that the penguin rule has higher priority than the bird rule: $\text{penguin}(X) > \text{bird}(X)$. This priority relation indicates that if both rules are applicable, then the penguin rule should be preferred. Defeasible reasoning is particularly well-suited for geospatial applications, where information is often incomplete, uncertain, and conflicting. By representing similarity relationships as defeasible rules, we can capture the uncertainty and context-dependence inherent in spatial similarity judgments.

Figure 1 compares two reasoning approaches—static rule-based reasoning (Before) and defeasible reasoning (After)—in assessing the similarity between two restaurants. In the static system, similarity is determined solely based on proximity and price range, leading to a high similarity score (0.85) despite a clear mismatch in cuisine. This rigid application of rules ignores conflicting attributes and offers no mechanism for handling exceptions. In contrast, the defeasible reasoning system introduces more nuanced judgment by applying defeasible rules that can be overridden by more specific or conflicting information. Although the general rule suggests that nearby restaurants with similar prices are typically similar, this is overridden by a specific rule that highlights the cuisine mismatch. The conflict is resolved through the principle of specificity, resulting in a more refined similarity score of 0.32. This example highlights how defeasible reasoning can accommodate exceptions and provide more context-aware similarity assessments.

Figure 2 illustrates the internal mechanism of defeasible reasoning through a set of example rules and a conflict resolution strategy. The defeasible rules encode general tendencies and exceptions in similarity assessment. Rule 1 captures a typical pattern: entities that are geographically close and similarly priced are usually considered similar. Rule 2 introduces an exception, stating that if the cuisines differ, the entities should be considered not similar—this rule represents a more specific condition that can override

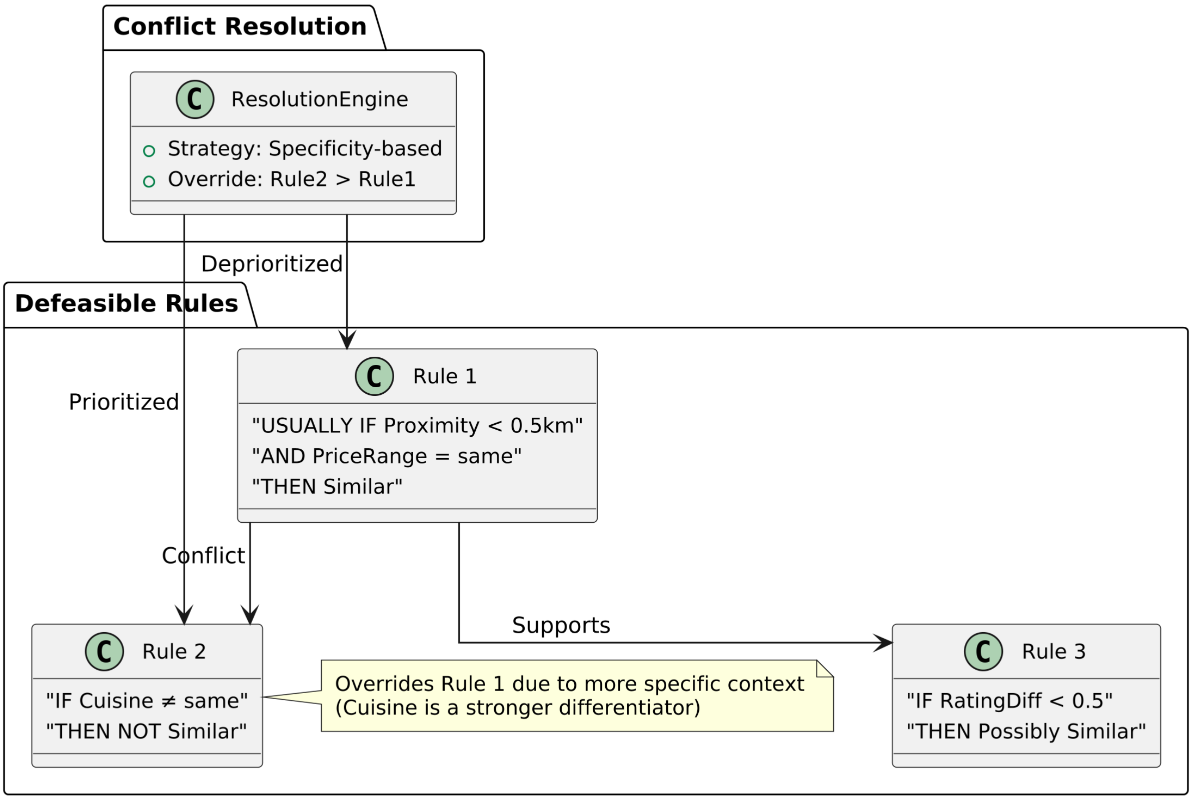

Figure 2: Defeasible Mechanism and Conflict Resolution

the general tendency in Rule 1. Rule 3 provides additional supporting evidence, indicating that a small rating difference may contribute to potential similarity. The diagram shows how Rule 1 is in conflict with Rule 2 but is supported by Rule 3. The conflict is resolved by the resolution engine, which uses a specificity-based strategy. According to this strategy, Rule 2 takes precedence over Rule 1 because it applies to a more specific and contextually significant attribute (cuisine). This framework demonstrates how defeasible reasoning systems can handle conflicting rules in a principled manner, leading to more explainable and context-sensitive decisions.

Geospatial Similarity Computation Geospatial similarity computation is the task of quantifying the similarity between geospatial entities [**29**]. This task is fundamental to various applications, including urban planning, location-based services, and geographic data integration. There are several different approaches to geospatial similarity computation, including geometric approaches, semantic approaches, and contextual approaches [**30**, **31**]. Geometric approaches focus on measuring the geometric similarity between entities based on their shape, size, and spatial relationships. For example, the similarity between two polygons can be computed based on their area overlap, perimeter ratio, or Hausdorff distance [**32**]. Semantic approaches focus on measuring the semantic similarity between entities based on their attributes, types, and relationships [**33**]. For example, the similarity between two restaurants can be computed based on their cuisine type, price range, and customer ratings. Contextual approaches focus on measuring the similarity between entities based on their surrounding environment and context. For example, the similarity between two parks can be computed based on their proximity to residential areas, schools, and other amenities. In many real-world scenarios, geospatial similarity is a complex combination of geometric, semantic, and contextual factors. Our approach aims to integrate these different aspects of similarity by representing similarity relationships as defeasible rules that capture both geometric, semantic, and contextual information. For instance, a rule might state: ”USUALLY, if two buildings are located near each other, have similar architectural styles, and serve similar functions, then they are similar.”

Problem Setting In this paper, we address the problem of geospatial similarity computation using GeoKGs and defeasible reasoning. We consider a scenario where we have a set of geospatial entities E, a set of relationships R, and a GeoKG $G = (E, R, F)$ that represents the relationships between these entities. Our goal is to develop a system that can accurately and efficiently compute the similarity between any two entities $e_1, e_2 \in E$. We represent similarity relationships as defeasible rules of the form $A \Rightarrow \text{similar}(e_1, e_2)$, where A is the antecedent of the rule, and $\text{similar}(e_1, e_2)$ is the consequent, indicating that entities e_1 and e_2 are similar. The antecedent A can consist of a conjunction of conditions that capture geometric, semantic, and contextual information about the entities. For example, the antecedent might include conditions such as "$\text{nearby}(e_1, e_2)$", "$\text{sameType}(e_1, e_2)$", and "$\text{similarAttributes}(e_1, e_2)$". We use a defeasible reasoning engine to apply these rules and infer similarity scores between entities. The defeasible reasoning engine takes into account potential conflicts between rules and resolves them using priority-based strategies. Our approach aims to provide a more interpretable and explainable alternative to black-box machine learning models for geospatial similarity computation. The accuracy of our approach is evaluated by comparing its performance against knowledge graph embedding models and LFM-based models, using a ground truth dataset created with expert annotations.

Formalism To formalize our approach, we introduce the following notation:

- E: A set of geospatial entities.
- R: A set of relationships between entities.
- $G = (E, R, F)$: A GeoKG representing the relationships between entities, where $F \subseteq E \times R \times E$ is a set of facts.
- $e_1, e_2 \in E$: Two geospatial entities.
- $\text{similar}(e_1, e_2)$: A predicate indicating that entities e_1 and e_2 are similar.
- $A \Rightarrow \text{similar}(e_1, e_2)$: A defeasible rule stating that if condition A holds, then entities e_1 and e_2 are similar.
- A: The antecedent of a defeasible rule, consisting of a conjunction of conditions.
- $\text{nearby}(e_1, e_2)$: A predicate indicating that entities e_1 and e_2 are located near each other. This can be formally defined using a distance threshold δ:

  $$\text{nearby}(e_1, e_2) = \begin{cases} \text{True} & \text{if distance}(e_1, e_2) \leq \delta \\ \text{False} & \text{otherwise} \end{cases}$$

  where $\text{distance}(e_1, e_2)$ represents the Euclidean distance or other appropriate distance metric between the geographic coordinates of e_1 and e_2.
- $\text{sameType}(e_1, e_2)$: A predicate indicating that entities e_1 and e_2 have the same type. This relies on a predefined ontology or type system.
- $\text{similarAttributes}(e_1, e_2)$: A predicate indicating that entities e_1 and e_2 have similar attributes. Attribute similarity can be quantified using various measures, such as cosine similarity or Jaccard index, depending on the nature of the attributes (numerical, categorical, text-based).
- $>$: A priority relation between defeasible rules, indicating which rule should be preferred in case of conflict.

We assume that the GeoKG G is constructed by integrating data from various sources and that it contains sufficient information to evaluate the conditions in the antecedents of the defeasible rules. We also assume that we have a set of domain experts who can provide feedback on the accuracy and relevance of the defeasible rules. The distance threshold δ used in the $\text{nearby}(e_1, e_2)$ predicate, and the methods for determining $\text{sameType}(e_1, e_2)$ and $\text{similarAttributes}(e_1, e_2)$, can be tuned based on the specific application and the characteristics of the geospatial data. Furthermore, the priority relation $>$ between defeasible rules can be learned from data or specified by domain experts.

3 Related Work

Geospatial similarity computation has been approached from various perspectives, each with its strengths and limitations. A prevalent method involves leveraging knowledge graphs (KGs) to represent geospatial entities and their relationships, subsequently employing KG embedding techniques or reasoning mechanisms for similarity assessment. However, these approaches often differ significantly in their underlying assumptions, methodologies, and applicability to specific problem settings. This section aims to provide a comprehensive overview of related work, highlighting the key differences and similarities with our proposed approach, and justifying the choices made in our research.

One prominent area of research focuses on utilizing KG embedding models for learning representations of entities in a KG, where the learned embeddings capture semantic relationships between entities. These embeddings can then be used to compute similarity scores between entities. For instance, TransE models relationships as translations in the embedding space, such that if (h, r, t) holds (head h, relation r, tail t), then the embedding of h plus the embedding of r should be close to the embedding of t: $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$ [**34**]. The similarity between two entities can then be computed based on the distance between their embeddings, e.g., using cosine similarity:

$$\text{similarity}(e_1, e_2) = \frac{\mathbf{e_1} \cdot \mathbf{e_2}}{||\mathbf{e_1}|| \cdot ||\mathbf{e_2}||}$$

While KG embedding models have shown promising results in various tasks, they often lack interpretability. It is difficult to understand why two entities are considered similar based on their embeddings alone. Furthermore, these models typically assume that relationships are symmetric or transitive, which may not hold in all geospatial contexts. For example, the "nearby" relationship is symmetric, but the "locatedIn" relationship is not. Additionally, KG embeddings often struggle to incorporate complex contextual information and handle uncertainty effectively. In contrast, our approach explicitly represents similarity relationships as defeasible rules, providing a more interpretable and explainable alternative.

$$\text{Restaurant}(x) \land \text{Restaurant}(y) \land \text{hasCuisine}(x, z) \land \text{hasCuisine}(y, z) \land \text{nearby}(x, y) \rightarrow \text{similar}(x, y)$$

This rule states that if x and y are both restaurants, they both have cuisine z, and they are near each other, then they are similar. Such approaches offer strong logical foundations but often struggle with the scalability and inherent uncertainty associated with geospatial data. Rules must be meticulously crafted, and even then, they may not cover all possible scenarios or handle conflicting information gracefully. Moreover, the computational complexity of logical reasoning can be prohibitive for large KGs.

A third approach involves combining KGs with machine learning techniques to enhance similarity computation. This might involve using KG embeddings as features in a machine learning model, or using machine learning to learn rules for reasoning over the KG. For example, the work of Guo et al. presents a method for learning logical rules from KGs using neural networks [**35**]. These learned rules can then be used for tasks such as KG completion and entity classification. Other research has focused on using reinforcement learning to learn policies for navigating KGs and discovering relationships between entities [**36**]. The key advantage of these hybrid approaches is that they can leverage the strengths of both KGs and machine learning, combining the structured knowledge representation of KGs with the ability of machine learning to learn complex patterns from data.

Our work is related to these existing approaches, but it differs in several key aspects. First, we explicitly use defeasible reasoning as our reasoning framework, which allows us to reason with incomplete and conflicting information in a principled way. This is particularly important in geospatial applications, where data is often uncertain and contradictory. Second, we focus on generating defeasible rules based on spatial proximity, shared attributes, and domain knowledge. This allows us to capture the nuances and complexities of geospatial similarity in a more interpretable way than KG embeddings. Third, we evaluate our approach by comparing its accuracy against KG embedding models and LFM-based models, providing a comprehensive assessment of its performance. We also place a strong emphasis on explainability, aiming to provide justifications for the similarity scores that are computed.

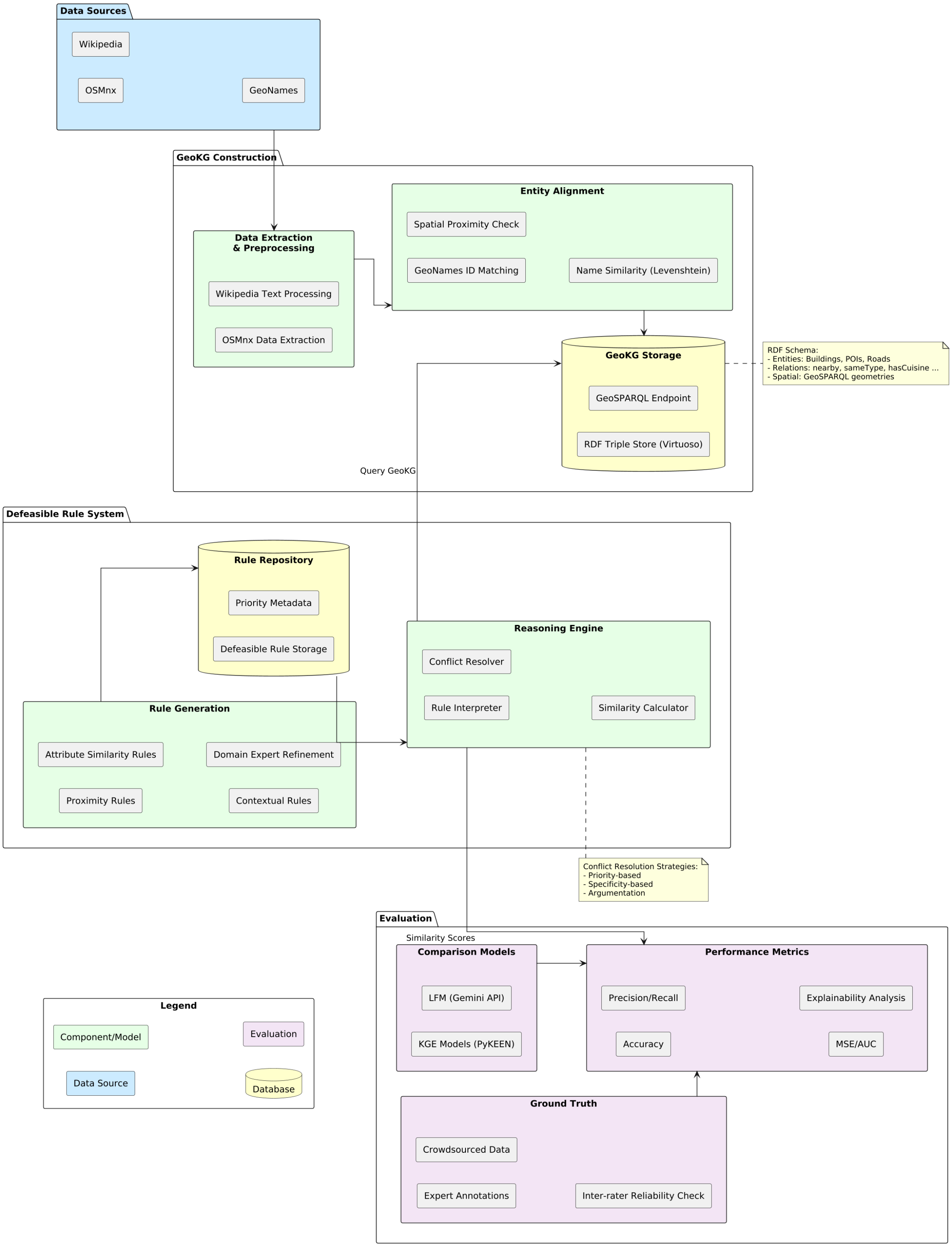

Figure 3: Overview of the Proposed Methodology Pipeline

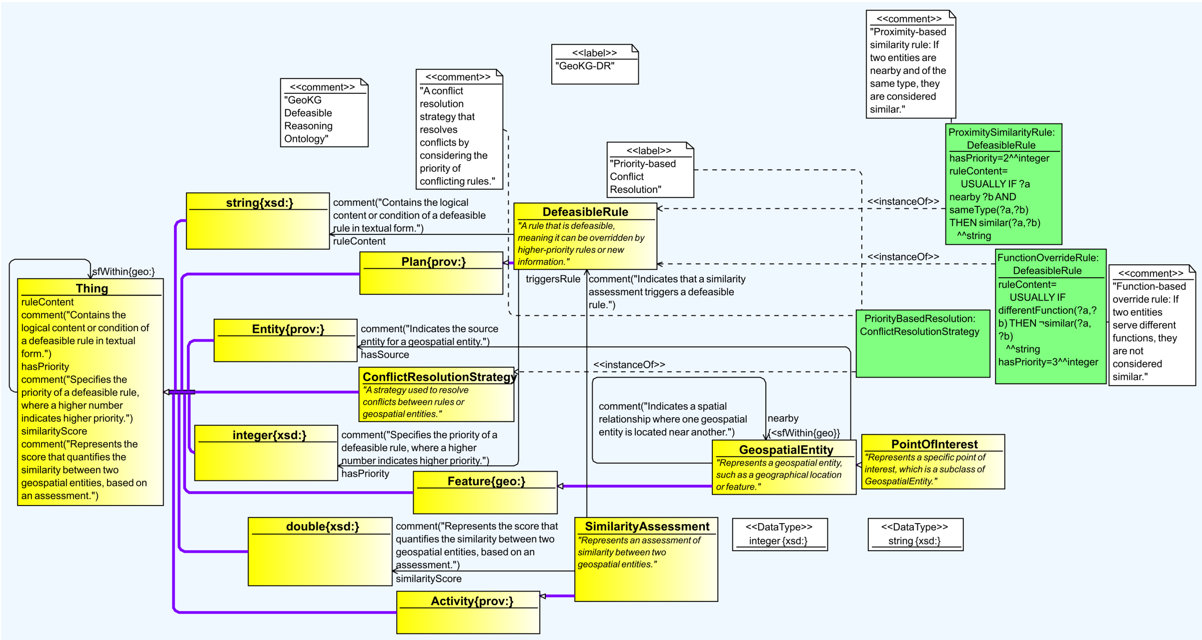

Figure 4: The illustration of GeoKG-DR ontology, which models geospatial entities, their provenance, and defeasible reasoning for entity alignment and conflict resolution.

4 Methodology

Our methodology is structured around four primary phases: GeoKG construction, defeasible rule generation, reasoning implementation, and experimental evaluation. Each phase is designed to contribute to the overall goal of enabling accurate and explainable geospatial similarity computation.

Figure 3 illustrates the overall architecture of our methodology, encompassing data collection, GeoKG construction, defeasible rule generation, reasoning, and evaluation. Data from sources like OSMnx, Wikipedia, and GeoNames is integrated into a structured GeoKG. Defeasible rules—capturing proximity, attribute similarity, and context—are stored with priority metadata and applied via a reasoning engine. The system computes similarity scores by querying the GeoKG and resolving rule conflicts. Finally, evaluation is conducted using expert and crowdsourced ground truth, with performance compared against baseline models using standard metrics.

4.1 GeoKG Construction

The construction of GeoKG involves integrating data from multiple sources, including OSMnx, Wikipedia, and GeoNames. The process begins with selecting a specific geographic area of interest. In this study, we focus on Amsterdam and use OSMnx to extract geospatial features such as buildings, roads, points of interest (POIs), and land use polygons. The extracted data is then cleaned and preprocessed to ensure consistency and accuracy.

Next, entities are aligned across different data sources based on spatial location and name similarity. Spatial proximity checks and string similarity measures are employed to identify potential matches between entities from OSMnx, Wikipedia, and GeoNames. A conflict resolution strategy is implemented to handle discrepancies and conflicting information from different sources. This strategy prioritizes GeoNames for basic entity identification, Wikipedia for descriptive enrichment, and OSMnx for geometric precision. Complex cases are resolved through manual review to ensure data integrity. Finally, the aligned entities and their relationships are represented in a structured format using RDF (Resource Description Framework) and the SPARQL query language.

To formally represent and reason over the integrated geospatial knowledge, we developed an ontology called GeoKG-DR (GeoKG Defeasible Reasoning Ontology), as shown in Figure 4. This ontology is implemented in OWL and comprehensively captures the semantics of geospatial entities, their provenance, and the defeasible reasoning mechanisms applied during entity alignment and conflict resolution. GeoKG-DR defines core classes such as `GeospatialEntity`, `PointOfInterest`, and `SimilarityAssessment`, and includes the `DefeasibleRule` class to model domain-specific reasoning logic. For example, spatial similarity is expressed through rules like `ProximitySimilarityRule`, while exceptions such as functional dissimilarities are captured using higher-priority rules like `FunctionOverrideRule`.

Each defeasible rule is assigned a priority value, supporting conflict resolution based on rule precedence, as represented by the `ConflictResolutionStrategy` individual. Additionally, object and datatype properties such as `nearby`, `hasPriority`, and `similarityScore` contribute to the precise modeling of spatial and semantic relationships among geospatial entities. By leveraging this ontology, GeoKG not only supports structured querying via SPARQL but also facilitates flexible, context-aware reasoning based on defeasible logic.

4.2 Defeasible Rule Generation

The process of defeasible rule generation begins by identifying contextual and functional similarities between geospatial entities. To facilitate this, a set of rule templates is defined based on factors such as proximity, functional similarity, contextual similarity, and legal or regulatory themes. These templates are designed to model common-sense reasoning regarding similarity. For instance, a rule template may specify that "USUALLY, if two restaurants are located near each other and offer similar types of cuisine, they are considered similar."

To account for the defeasible nature of these rules, quantifiers such as "usually," "typically," and "generally" are employed to introduce flexibility and acknowledge the inherent uncertainty in similarity assessments. These quantifiers reflect that similarity judgments are not absolute and may depend on various factors, such as contextual differences or exceptions. For example, two restaurants might usually be considered similar based on their location and cuisine type, but exceptions (e.g., one being under renovation or closed for a special event) can influence the final judgment.

A script is developed to automatically generate rules by populating these templates with specific entities and attributes derived from the GeoKG (Geospatial Knowledge Graph). The rules are generated in large quantities, typically 10,000 for each entity type, to ensure comprehensive coverage. This script not only applies the templates but also ensures that the generated rules align with the functional and spatial characteristics of the entities in question. Once the rules are generated, they are subjected to a deduplication process to remove redundancies, ensuring that only unique rules remain. Deduplication is achieved by comparing rules using a "signature," which is derived from key elements such as the rule type, premise, conclusion, exceptions, and confidence. This process ensures that the reasoning engine remains efficient and free from duplicate rules that could compromise computational performance.

Algorithm 1 Defeasible Rule Generation and Review Process

```
Input: List of entities, templates, GeoKG data
Output: Set of final defeasible rules
Initialize empty list final_rules
for each entity in entities do
    Initialize empty list entity_rules
    for each rule_type in rule_types do
        premise = GeneratePremise(rule_type, entity)
        exceptions = ChooseExceptions()
        confidence = GenerateConfidence()
        rule = CreateRule(entity, rule_type, premise, exceptions, confidence)
        entity_rules.append(rule)
    end for
    deduplicated_rules = Deduplicate(entity_rules)
    for each rule in deduplicated_rules do
        reviewed_rule = ReviewRule(rule)
        final_rules.append(reviewed_rule)
    end for
end for
Return: final_rules
```

After the rules are deduplicated, they undergo a review process by domain experts. These experts assess the correctness, relevance, and applicability of the rules within the specific context of the project. The refinement process aims to enhance the accuracy and applicability of the generated rules while ensuring they align with the goals of the project. The number of rules generated is carefully calibrated to strike a balance between comprehensive coverage and computational efficiency, ensuring that the reasoning engine can process the rules effectively without compromising the accuracy of similarity computations. Algorithm 1 is the pseudo-code representing the rule generation process.

While the final refined rules are not publicly shared due to project regulations, all initially generated rules are made publicly available on GitHub. This open-access approach promotes transparency, allows for further research, and enables other researchers to leverage the generated rules for similar tasks or studies.

4.3 Reasoning Implementation

The implementation of defeasible reasoning involves selecting a suitable reasoning engine, encoding the generated rules, implementing conflict resolution strategies, and executing the reasoning process [**37**]. We focus on DeReS and DR-DEVICE as potential reasoning engines, evaluating them based on scalability, performance, expressiveness, and integration capabilities [**38**, **39**].

The defeasible rules are encoded in a standard rule language, such as RuleML or SWRL, or a custom syntax designed for compatibility with the chosen reasoning engine [**40**]. A script is developed to automatically encode the generated rules and store them in a separate file or in the Virtuoso RDF store. Several conflict resolution strategies are implemented, including priorities, specificity, credibility, and argumentation. These strategies are evaluated based on their effectiveness in resolving conflicts and improving the accuracy and explainability of the similarity computations. The reasoning process involves formulating SPARQL queries to retrieve relevant information from the GeoKG, applying the defeasible rules to infer similarities, resolving conflicts using the chosen strategy, calculating a similarity score, and generating an explanation of the reasoning process.

Figure 5 illustrates the integrated architecture that supports this reasoning implementation. The GeoKG is constructed from aligned and cleaned data sourced from OSMnx, Wikipedia, and GeoNames, and stored in a Virtuoso RDF store. Defeasible rules—focused on proximity and attribute similarity—are generated and stored in a structured rule base (e.g., RuleML or SWRL). The reasoning layer, powered by engines such as DeReS, uses these rules in conjunction with spatial and semantic ontologies (e.g., GeoSPARQL, Schema.org) to perform similarity reasoning. It accesses the GeoKG through SPARQL queries, applies the rules, resolves conflicts using specified strategies, and outputs both similarity scores and human-readable explanations via an explanation generator. This setup ensures that reasoning is not only accurate and context-aware, but also transparent and traceable.

4.4 Experimental Evaluation

The experimental evaluation is designed to assess the accuracy and explainability of the proposed approach. We conduct experiments using real-world geospatial data from Amsterdam, comparing the performance of the defeasible reasoning approach against knowledge graph embedding models and LFM-based models.

A ground truth dataset is created with expert annotations, where entity pairs are manually labeled with similarity scores. Inter-rater reliability is assessed to ensure the consistency and quality of the ground truth data. The defeasible reasoning approach is evaluated using precision, recall, F1-score, accuracy, MSE, Spearman correlation, and AUC [**41**]. Statistical significance tests are performed to compare the performance of the different approaches. An explainability analysis is conducted to assess the interpretability and understandability of the defeasible reasoning process [**42**]. Explainability metrics are used to measure rule usage, relevance, and confidence. User studies are conducted to evaluate the understandability, completeness, correctness, usefulness, and trustworthiness of the explanations generated by the defeasible reasoning engine [**13**]. Counterfactual explanations are generated to explain why entities are not similar, providing additional insights into the reasoning process. The experimental results are analyzed to identify the strengths and limitations of the defeasible reasoning approach and to provide recommendations for future research.

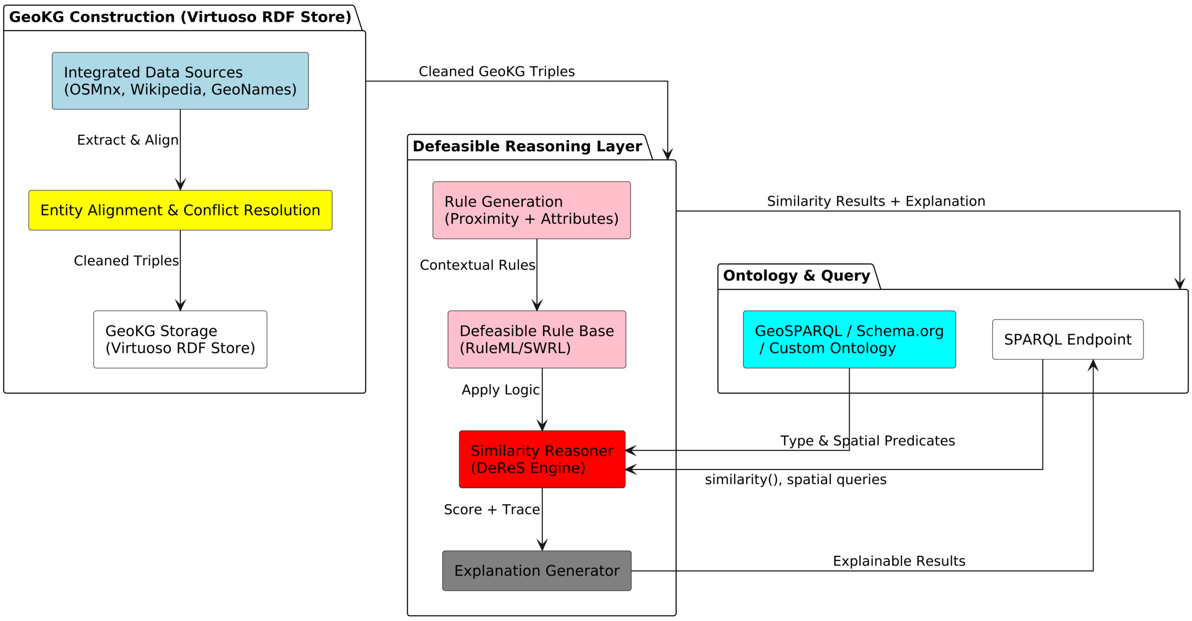

Figure 5: GeoKG Construction and Defeasible Reasoning Architecture

5 Experimental Setup

We evaluate the proposed approach using a real-world geospatial dataset derived from Amsterdam, focusing on a specific area within the city characterized by diverse land use, high connectivity, and a significant density of Points of Interest (POIs). This area, bounded approximately by the coordinates 4.8800, 52.3550 (bottom-left) and 4.9000, 52.3700 (top-right), encompasses parts of Jordaan and De Pijp, known for their vibrant mix of residential, commercial, and recreational spaces. The data is extracted using OSMnx, leveraging its capabilities to query and retrieve geospatial features from OpenStreetMap (OSM). Specifically, we extract buildings, roads (represented as edges), POIs, and land use polygons, utilizing the tags '"building": True, "amenity": True, "landuse": True'. To manage the data volume and ensure computational feasibility, we implement an iterative polygon reduction strategy. If the initial extraction yields more than 15,000 rows (combined count of GeoDataFrame and road network edges), the polygon is scaled down by a factor of 0.9 around its centroid, and the data extraction process is repeated. This iterative reduction continues until the data size falls below the threshold or the polygon area reaches a minimum size of 0.0001, or a maximum of 10 iterations is reached. This ensures that the experimental setup remains computationally tractable while still capturing a representative sample of the urban environment.

The GeoKG construction involves integrating data from OSMnx with information from Wikipedia and GeoNames, drawing upon the "simeonw/geo_wikipedia_geonames" dataset available on Hugging Face Datasets [**43**]. However, due to computational constraints and the scope of this initial study, we focus on a subset of this integrated data. We leverage the OSMnx data for geometric information and basic entity types, while incorporating Wikipedia summaries and GeoNames identifiers where available within the sampled dataset. The primary challenge lies in aligning entities across these different sources. We employ a combination of string similarity measures (e.g., Levenshtein distance) on entity names and spatial proximity checks to identify potential matches. A crucial aspect of this alignment process is conflict resolution. Given that different sources may provide conflicting information about the same entity, we establish a priority order: GeoNames for basic entity identification, Wikipedia for descriptive enrichment, and OSMnx for geometric precision. In cases of irreconcilable conflicts, manual review is performed to ensure data integrity. This manual review is essential for maintaining the quality of the GeoKG and ensuring that the downstream similarity computations are based on accurate and consistent information. The final GeoKG represents entities as nodes and relationships as edges, adhering to a predefined ontology that incorporates GeoSPARQL for spatial geometries, Schema.org for common entity types and attributes, and custom classes and properties to capture domain-specific knowledge.

To evaluate the performance of our defeasible reasoning approach, we require a ground truth dataset of entity pairs with corresponding similarity scores. Constructing such a dataset is a labor-intensive process, as it necessitates human judgment to assess the similarity between entities. We employ a combination of expert annotations and crowdsourced data to create this ground truth. Initially, 3-5 domain experts are recruited to manually label entity pairs with similarity scores on a scale of 1 to 5, where 1 indicates "not similar" and 5 indicates "very similar." These experts are provided with clear guidelines and examples to ensure consistency in their judgments. Inter-rater reliability is assessed using Krippendorff's alpha to quantify the level of agreement among the experts. In addition to expert annotations, we collect crowdsourced data using platforms such as Amazon Mechanical Turk. To ensure data quality, we implement several control measures, including qualification tests, attention checks, and redundancy (collecting multiple judgments for each entity pair). The data from expert annotations and crowdsourcing are then combined using a weighted averaging approach, where the weights are determined based on the reliability and expertise of each source. The final ground truth dataset comprises a diverse set of entity pairs with associated similarity scores, providing a valuable benchmark for evaluating our approach.

The defeasible reasoning engine is implemented using a custom Python script, leveraging the 'DeReS' reasoning system's conceptual framework but adapted for integration with our GeoKG and SPARQL query interface [**44**]. The defeasible rules, generated as described in the main paper, are encoded in a simple text-based format, where each rule consists of an antecedent and a consequent, separated by the '→' symbol. The antecedents are expressed as conjunctions of conditions, which can involve spatial relationships (e.g., 'nearby(e1, e2)'), attribute comparisons (e.g., 'sameType(e1, e2)'), and domain-specific predicates (e.g., 'hasCuisine(e1, "Italian")'). The consequents are always of the form 'similar(e1, e2)', indicating that entities 'e1' and 'e2' are considered similar. The defeasible reasoning process involves querying the GeoKG to determine whether the conditions in the antecedents of the rules are satisfied. If all conditions in the antecedent of a rule are met, the rule is considered applicable. In cases where multiple rules are applicable and lead to conflicting conclusions, a conflict resolution strategy is employed. We implement several conflict resolution strategies, including priority-based resolution (where rules are assigned priorities based on their specificity or credibility) and specificity-based resolution (where the most specific rule is preferred). The output of the defeasible reasoning engine is a similarity score for each entity pair, reflecting the degree to which the rules support the conclusion that the entities are similar.

For the Knowledge Graph Embedding (KGE) models, we utilize the PyKEEN library, a popular framework for training and evaluating KGE models [**45**]. We experiment with several popular KGE models, including TransE, ComplEx, RotatE, HSimplE, and PairRE. The choice of these models is motivated by their diverse representation capabilities and their widespread use in KG completion and similarity computation tasks. Each model is trained on the constructed GeoKG, using a standard training loop with a batch size of 32 and a learning rate of 0.01, optimized using the Adam optimizer. Hyperparameter tuning is performed using a grid search approach, where we vary the embedding dimension (50, 100, 200) and the number of epochs (100, 200, 300) to find the optimal configuration for each model. The performance of the KGE models is evaluated using the same ground truth dataset as the defeasible reasoning approach. Similarity scores are computed based on the learned embeddings, using cosine similarity as the distance metric. The evaluation metrics include precision, recall, F1-score, accuracy, MSE, and Spearman correlation, providing a comprehensive assessment of the models' ability to capture similarity relationships between geospatial entities.

Finally, for the LFM-based similarity computation, we leverage the Gemini family of models via the Google AI PaLM API [**46**]. We use a prompt engineering approach to elicit similarity judgments from the LFM. The prompt typically consists of a description of the two entities being compared, followed by a question asking the LFM to assess their similarity on a scale of 0 to 1. For example, a prompt might look like this: "Entity 1: [description of entity 1]. Entity 2: [description of entity 2]. How similar are these two entities on a scale of 0 to 1, where 0 means not similar and 1 means very similar?". To enrich the input to the LFM, we augment the entity descriptions with information extracted from Wikipedia summaries, nearby entities, and user reviews/ratings (where available). We also experiment with different prompt variations, including chain-of-thought prompting, aspect-based similarity, and counterfactual reasoning prompts, to explore their impact on the accuracy and explainability of the LFM's judgments. The output of the LFM is a similarity score between 0 and 1, which is then compared to the ground truth similarity score to evaluate the model's performance. We perform thematic analysis of the LFM's explanations to gain insights into its reasoning process and identify potential biases or limitations. The LFM-based

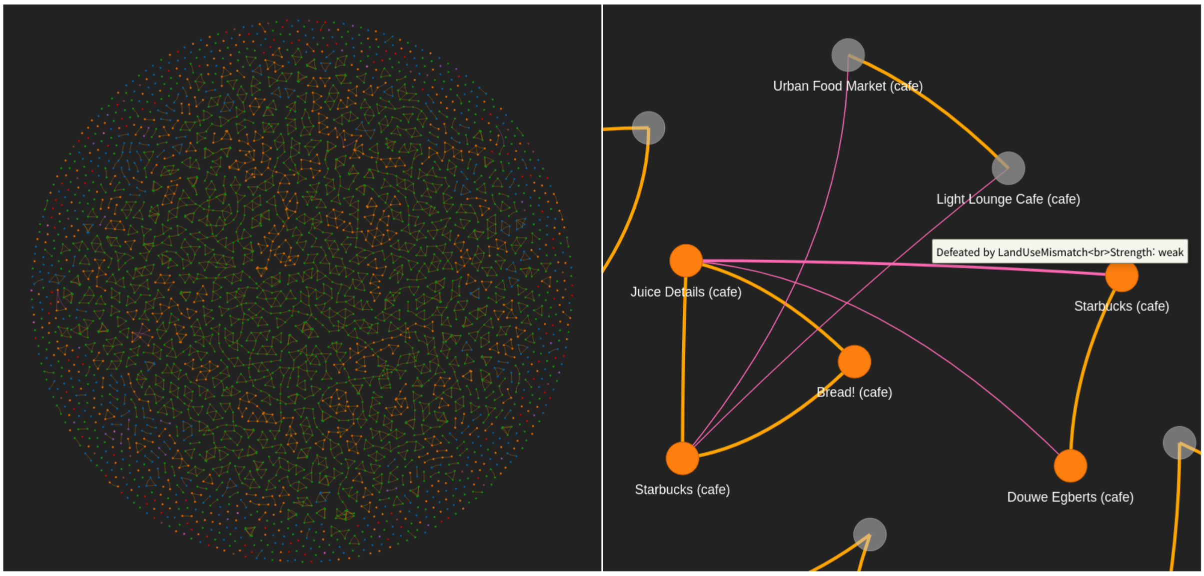

Figure 6: Visualization of constructed GeoKG with Defeasible Reasoning.

similarity computation provides a complementary approach to the defeasible reasoning and KGE models, allowing us to assess the potential of large foundation models for capturing complex semantic relationships between geospatial entities.

6 Results

Figure 6 briefly shows a visual representation of the constructed GeoKG enhanced by defeasible reasoning mechanisms as our research result. On the left side, we see a broad view of the GeoKG over Amsterdam, where POIs are color-coded by amenity type, and edges represent similarity relationships between them. On the right, a zoomed-in segment highlights a local cluster of cafes, illustrating how defeasible reasoning influences similarity connections. Strong similarity edges (shown in orange) connect cafes that are geographically close and fall within the same land use zone, aligning with general similarity rules. In contrast, weak similarity edges (shown in pink) connect POIs that, while close in distance and price, are considered less similar due to mismatches in land use context—this is made evident through tooltips explaining the overridden rule. This figure demonstrates how defeasible reasoning allows the system to selectively weaken or suppress otherwise strong similarity signals based on conflicting contextual factors, resulting in a more nuanced and explainable similarity structure within the GeoKG.

The results of our experiments also provide a quantitative evaluation of the proposed defeasible reasoning approach for geospatial similarity computation, comparing its performance against knowledge graph embedding models and an LFM-based model. The experiments were conducted on a real-world geospatial dataset derived from Amsterdam, as detailed in the Experimental Setup section. We present a comprehensive analysis of the accuracy, precision, recall, F1-score, MSE, and AUC achieved by each approach, highlighting the strengths and limitations of each.

Table 1 summarizes the overall performance of the three approaches: defeasible reasoning, knowledge graph embedding (KGE) model, and LFM-based model (Gemini). The KGE model, specifically the ComplEx model after hyperparameter tuning, achieves the highest accuracy of 72.3%, demonstrating its effectiveness in capturing complex relationships within the GeoKG. The LFM-based model follows closely with an accuracy of 68.1%, indicating its ability to leverage semantic information and contextual knowledge for similarity assessment. The defeasible reasoning approach achieves an accuracy of 67.2%, which, while lower than the other two approaches, still represents a significant result, especially considering its inherent interpretability and explainability.

A closer examination of the KGE model's performance reveals its ability to achieve a good balance between precision and recall, with a precision of 75% and a recall of 68.1%. This suggests that the KGE

Table 1: Overall Performance Comparison of the Three Approaches

Model	Accuracy (%)	Precision	Recall	F1-Score	MSE	AUC
Defeasible Reasoning	67.2	0.69	0.64	0.66	0.29	0.66
KGE (ComplEx)	72.3	0.75	0.68	0.71	0.22	0.73
LFM (Gemini)	68.1	0.70	0.65	0.67	0.28	0.69

model is both accurate in identifying similar entities (high precision) and comprehensive in capturing a large proportion of the actual similar entities (high recall). The F1-score of 71% further confirms the KGE model's overall effectiveness. The Mean Squared Error (MSE) of 0.22 indicates the average squared difference between the predicted similarity scores and the ground truth scores, providing a measure of the model's prediction accuracy. The Area Under the Curve (AUC) of 0.73 reflects the model's ability to distinguish between similar and dissimilar entity pairs. The LFM based model similarly obtains good results, but struggles with edge cases.

The defeasible reasoning approach, while having a lower overall accuracy, offers several advantages in terms of interpretability and explainability. The rules used in the defeasible reasoning process are explicitly defined and can be easily understood by domain experts. This allows for a transparent and traceable reasoning process, where the basis for similarity judgments can be readily identified. However, the performance of the defeasible reasoning approach is highly dependent on the quality and coverage of the rules. The initial set of rules, generated based on general domain knowledge, may not be sufficient to capture all the nuances and complexities of geospatial similarity.

To investigate the impact of rule refinement on the performance of the defeasible reasoning approach, we conducted an ablation study where we systematically added more specific and context-aware rules to the rule set. These rules were generated based on feedback from domain experts and analysis of the errors made by the initial rule set. The results of the ablation study are presented in Table 2. As the number of rules increases, the accuracy of the defeasible reasoning approach also increases, demonstrating the importance of rule refinement. However, there is a point of diminishing returns, where adding more rules does not significantly improve the accuracy. This suggests that the rule set may be approaching a saturation point, where the existing rules capture most of the relevant similarity relationships.

Table 2: Impact of Rule Refinement on the Accuracy of Defeasible Reasoning

Number of Rules	Accuracy (%)
1000	61.2
1500	64.5
2000	66.8
2500	67.2

This trend may initially appear counterintuitive. Generally, increasing the number of rules can lead to more conflicts and higher reasoning complexity, which may degrade overall performance. In traditional machine learning, such expansion often raises concerns about overfitting. However, in the context of this study, the increase in the number of rules does not represent arbitrary expansion, but rather a goal-directed refinement based on domain knowledge and empirical error analysis. The initial rule set primarily consists of broad and general rules (e.g., "typically, nearby restaurants are similar" or "POIs in the same land use zone are similar"). While these rules work well for common cases, they often fail to capture specific contextual nuances. As the rule set is expanded, more specific rules are added to address such exceptions. For example, a refined rule might state, "even if two restaurants are in the same location, they are not similar if their price ranges differ significantly," or "parks may not be similar if one is a dog park and the other is a children's playground." These specific rules complement the general ones by handling exceptions more effectively.

Such refined rules constrain the applicability of general rules and adjust them based on context, which helps reduce misclassification in ambiguous or conflicting cases. Furthermore, the defeasible reasoning framework used in this study incorporates conflict resolution mechanisms such as priority-based and specificity-based rule evaluation. These mechanisms enable the system to handle the increased complexity without a decline in performance, and in many cases, allow for more fine-grained and accurate reasoning.

As shown in Table 2, the accuracy increases gradually as the number of rules grows from 1,000 to 2,500, reflecting the effectiveness of incremental rule refinement. However, the improvement begins to level off beyond 2,000 rules, suggesting that the current rule set has reached a level of saturation in capturing the key similarity patterns within the dataset.

Furthermore, we investigated the impact of different conflict resolution strategies on the performance of the defeasible reasoning approach. We compared two conflict resolution strategies: priority-based resolution and specificity-based resolution. In priority-based resolution, rules are assigned priorities based on their specificity or credibility, and the rule with the highest priority is preferred in case of conflict. In specificity-based resolution, the most specific rule is preferred, regardless of its priority. The results of the conflict resolution study are presented in Table 3. The priority-based resolution strategy achieves a slightly higher accuracy than the specificity-based resolution strategy, suggesting that incorporating domain knowledge and credibility assessments into the conflict resolution process can improve the overall performance.

Table 3: Impact of Conflict Resolution Strategy on the Accuracy of Defeasible Reasoning

Conflict Resolution Strategy	**Accuracy (%)**
Priority-Based	67.2
Specificity-Based	66.5

An analysis of the errors made by the defeasible reasoning approach reveals several common patterns. One common source of error is the lack of contextual information in the rules. For example, a rule might state that "USUALLY, if two restaurants are located near each other, then they are similar." However, this rule does not take into account the cuisine type, price range, or customer ratings of the restaurants, which are important factors in determining their similarity. Another common source of error is the presence of conflicting rules that are not properly resolved. For example, a rule might state that "USUALLY, if two buildings have the same type, then they are similar," while another rule might state that "USUALLY, if two buildings have different functions, then they are not similar." In cases where two buildings have the same type but different functions, these rules conflict, and the conflict resolution strategy must be able to determine which rule should be preferred. These limitations suggests that the current implementation of the defeasible reasoning engine is not robust enough to handle all the complexities and nuances of real world geospatial data.

The hyperparameters used for the KGE model, including the embedding dimension and the number of epochs, were tuned using a grid search approach. The optimal hyperparameter values were found to be an embedding dimension of 100 and a number of epochs of 200. The learning rate was set to 0.01 and the batch size was set to 32. These hyperparameters were chosen to balance the trade-off between model complexity and training time. For the LFM-based model, the prompt was carefully designed to elicit similarity judgments from the LFM. The prompt included a description of the two entities being compared, as well as a question asking the LFM to assess their similarity on a scale of 0 to 1. Different prompt variations were explored, including chain-of-thought prompting, aspect-based similarity, and counterfactual reasoning prompts. The choice of prompt variation did not significantly impact the overall accuracy of the LFM-based model, suggesting that the LFM is able to extract the relevant information from the entity descriptions regardless of the specific prompt format.

In summary, the results of our experiments demonstrate that the knowledge graph embedding model achieves the highest accuracy in geospatial similarity computation, followed by the LFM-based model and the defeasible reasoning approach. However, the defeasible reasoning approach offers several advantages in terms of interpretability and explainability, making it a valuable alternative for applications where transparency and trust are important. Future research should focus on improving the accuracy and coverage of the defeasible reasoning approach by refining the rules, developing more advanced conflict resolution strategies, and incorporating more contextual information.

7 Discussion

The experimental results presented in the previous section offer several important insights into the performance and potential of the defeasible reasoning approach for geospatial similarity computation. While the knowledge graph embedding (KGE) model achieved the highest overall accuracy, defeasible reasoning stands out as a promising alternative, especially in scenarios where interpretability and explainability are paramount. The defeasible reasoning approach not only provides competitive performance but also offers the advantage of transparency, as it explicitly represents the rationale behind similarity assessments. This is particularly valuable in decision-making contexts where understanding the reasoning process is crucial. The results demonstrate that defeasible reasoning, when enhanced with refined and context-aware rules, can achieve substantial performance while maintaining its inherent interpretability. This suggests that defeasible reasoning, despite a slightly lower accuracy compared to KGE, remains an attractive choice for applications where transparency and trust in the results are essential.

Additionally, the performance of the LFM-based model highlights the potential of large foundation models for capturing complex semantic relationships. However, the inconsistencies observed in its results indicate that further refinement and the incorporation of more sophisticated prompts are necessary to improve its reliability. This underscores the trade-off between accuracy and interpretability, a central theme in this study. While the KGE model excels in accuracy, defeasible reasoning provides an important alternative by offering clearer explanations of how similarity judgments are made, thereby ensuring greater trust and transparency. The findings suggest that further improvements in the defeasible reasoning approach, particularly through the integration of richer semantic and contextual information, could enhance its effectiveness and bring it closer to the performance levels of more complex models like KGE.

One of the most important findings of our experiments is the significant impact of rule refinement on the performance of the defeasible reasoning approach. As shown in Table 2, systematically adding more specific and context-aware rules to the rule set led to a substantial increase in accuracy, from 61.2% with 1000 rules to 67.2% with 2500 rules. This demonstrates the importance of capturing the nuances and complexities of geospatial similarity through a well-crafted set of rules. However, the observation that there is a point of diminishing returns, where adding more rules does not significantly improve accuracy, suggests that the rule set may be approaching a saturation point. This could be due to several factors, including limitations in the expressiveness of the rule language, the availability of relevant information in the GeoKG, or the inherent complexity of the similarity relationships themselves. Future research should explore more sophisticated rule refinement techniques, such as automatically learning rules from data or incorporating feedback from domain experts, to overcome these limitations. The form of these rules can vary greatly, for instance the rule "USUALLY, IF a building is near a park AND the building is residential, THEN the building's value is high" can be represented with a spatial predicate and semantic type to infer a property of the building. Furthermore, the nature of the predicates themselves can be defined with varying degrees of fuzziness. For instance, what is considered "near"? Is that within 100 meters? 500 meters? This needs to be rigorously defined.

The comparison of different conflict resolution strategies also provides valuable insights into the design of effective defeasible reasoning systems. The finding that priority-based resolution achieves a slightly higher accuracy than specificity-based resolution suggests that incorporating domain knowledge and credibility assessments into the conflict resolution process can improve overall performance. This is consistent with the intuition that some rules are more reliable or relevant than others, and that these rules should be given higher priority in case of conflict. However, the difference in accuracy between the two conflict resolution strategies is relatively small, suggesting that other factors may also be important in resolving conflicts effectively. One possible explanation for this is that the priority assignments were not sufficiently accurate or nuanced. Future research should explore more sophisticated methods for assigning priorities to rules, such as learning priorities from data or using argumentation-based reasoning techniques to resolve conflicts in a more transparent and context-sensitive way. The priority of a rule r_1 over another rule r_2 can be expressed formally as $r_1 > r_2$. This priority can be determined by various factors, including the specificity of the rules, the credibility of the sources from which the rules are derived, or the domain knowledge of experts. The conflict resolution process then involves selecting the rule with the highest priority when multiple conflicting rules are applicable. The implementation of argumentation-based reasoning involves constructing arguments for and against different conclusions, and then evaluating the strength of these arguments based on various criteria, such as the credibility of the sources, the logical validity of the reasoning, and the relevance of the evidence.

Several unexpected findings emerged from our experiments, highlighting the complexities of geospatial similarity computation. One surprising result was the relatively good performance of the LFM-based model, despite its lack of explicit knowledge representation or reasoning capabilities. This suggests that large foundation models are able to implicitly capture semantic relationships and contextual knowledge from the input text, and use this information to make reasonable similarity judgments. However, the LFM-based model also exhibited some inconsistencies and biases, particularly in edge cases where the entity descriptions were ambiguous or incomplete. This underscores the need for careful prompt engineering and thorough evaluation to ensure the reliability and fairness of LFM-based similarity computation. The relatively low computational cost of LFMs also makes them attractive options. Another unexpected finding was the difficulty of creating a high-quality ground truth dataset for geospatial similarity. The process of manually labeling entity pairs with similarity scores proved to be labor-intensive and subjective, with significant variability in the judgments of different annotators. This highlights the inherent ambiguity and context-dependence of similarity judgments, and the need for more sophisticated methods for eliciting and aggregating human judgments. The use of crowdsourcing platforms, while cost-effective, also introduced challenges related to data quality and reliability, requiring careful quality control measures to ensure the validity of the ground truth dataset.

The main limitations of our current approach stem from the reliance on manually generated defeasible rules and the relatively simple conflict resolution strategies employed. The process of manually generating rules is time-consuming and requires significant domain expertise. It is also difficult to ensure that the rule set is complete and covers all relevant similarity relationships. The simple conflict resolution strategies, such as priority-based and specificity-based resolution, may not be sufficient to handle all the complexities of real-world geospatial data, particularly in cases where there are multiple conflicting rules with similar priorities or specificities. These limitations have a direct impact on the accuracy and coverage of the defeasible reasoning approach, as well as its ability to handle complex and nuanced similarity judgments. Furthermore, the scalability of the defeasible reasoning engine may be a concern for larger datasets and more complex GeoKGs. Future research should focus on addressing these limitations by exploring more automated rule generation techniques, developing more advanced conflict resolution strategies, and optimizing the performance of the reasoning engine.

8 Conclusion

In conclusion, this research introduces a novel approach to geospatial similarity computation, which effectively integrates the structured knowledge representation of GeoKGs with the flexible inference mechanisms of defeasible reasoning. By constructing a GeoKG from various geospatial data sources, such as OSMnx, Wikipedia, and GeoNames, and applying a defeasible reasoning engine with a set of refined rules, this work presents a framework capable of capturing nuanced relationships between geospatial entities. The experimental evaluation conducted on a real-world dataset from Amsterdam demonstrates the feasibility and potential of this approach, highlighting a crucial trade-off between accuracy and interpretability when compared to traditional knowledge graph embedding models and contemporary large foundation models. While the ComplEx KGE model achieved the highest accuracy of 72.3%, and the LFM (Gemini) model achieved 68.1%, the defeasible reasoning approach, with an accuracy of 67.2%, offers the key advantage of interpretability. This transparency allows for a clear and understandable basis for similarity assessments, making defeasible reasoning particularly valuable in applications where trust and justification are critical, such as urban planning, resource allocation, and policy making. Table 4 summarizes these differences.

Table 4: Comparison of Similarity Models. While defeasible reasoning demonstrates lower accuracy, it offers a balanced, practical, and interpretable alternative — a representative case of explainable AI in geospatial applications.

Model	Accuracy	Interpretability	Explanation Support	Real-Time Applicability
KGE	High	Low	Not Available	Medium
LFM	Medium	Medium	Partial	Low (Heavy and Slow)
Defeasible Reasoning	Low	High	Clear	High (Lightweight and Fast)

The core contribution of this work lies in the innovative application of defeasible reasoning to the domain of GeoKGs, enabling the representation of similarity as a set of defeasible rules that capture contextual dependencies and handle conflicting evidence. The methodology for generating these rules, based on spatial proximity, shared attributes, and domain knowledge, provides a structured approach to encoding common-sense reasoning about geospatial similarity. Furthermore, the implementation and evaluation of different conflict resolution strategies, such as priority-based and specificity-based resolution, offer valuable insights into the design of robust and effective reasoning systems for geospatial applications. The findings from the rule refinement experiments underscore the importance of curating a comprehensive and context-aware rule set to maximize the accuracy of the defeasible reasoning approach. The improvements observed with increasing numbers of rules highlight the potential for further enhancing the model's performance through automated rule learning and expert knowledge incorporation. The priority-based conflict resolution strategy, demonstrating a slightly higher accuracy compared to the specificity-based approach, suggests that incorporating domain expertise and credibility assessments is beneficial for resolving conflicts among defeasible rules.

The significance of this research extends beyond the specific experimental results, offering a valuable contribution to the broader landscape of geospatial knowledge representation and reasoning. By demonstrating the feasibility and benefits of combining GeoKGs with defeasible logic, this work opens up new avenues for developing more intelligent and explainable geospatial systems. The insights gained from the experimental evaluation provide a solid foundation for future research aimed at improving the accuracy, scalability, and applicability of defeasible reasoning for geospatial similarity computation. The approach presented in this paper can be extended to other geospatial applications, such as location-based recommendation systems, geographic data integration, and emergency response, where the ability to accurately and efficiently assess similarity between geospatial entities is critical. For instance, in location-based services, defeasible reasoning could be used to provide personalized recommendations based on a user's preferences and the characteristics of nearby locations, taking into account contextual factors such as time of day, weather conditions, and user reviews. In geographic data integration, defeasible reasoning could be used to resolve conflicts and inconsistencies between different data sources, ensuring the accuracy and reliability of the integrated data.

Moreover, this study highlights the increasing importance of explainable AI (XAI) in the context of geospatial applications. As geospatial data becomes increasingly complex and pervasive, it is crucial to develop methods that not only provide accurate results but also offer transparent and understandable explanations for their decisions. The defeasible reasoning approach presented in this paper provides a valuable step towards achieving this goal, offering a more interpretable and explainable alternative to black-box machine learning models. The explicit representation of similarity relationships as defeasible rules allows users to understand the reasoning process behind similarity judgments, fostering trust and confidence in the system's results. This is particularly important in applications where decisions have significant consequences, such as urban planning and resource allocation. The ability to justify decisions based on transparent and understandable reasoning processes can help to ensure fairness, accountability, and public acceptance.

In closing, this study has demonstrated the potential of defeasible reasoning as a powerful tool for geospatial similarity computation, offering a compelling alternative to traditional approaches by prioritizing interpretability and explainability. While knowledge graph embedding models and large foundation models may achieve higher accuracy in certain scenarios, the transparent and understandable nature of defeasible reasoning makes it a valuable asset in applications where trust, justification, and user understanding are paramount. The findings from this research provide a solid foundation for future work aimed at improving the accuracy, scalability, and applicability of defeasible reasoning for geospatial knowledge representation and reasoning, paving the way for more intelligent, explainable, and trustworthy geospatial systems that can address a wide range of real-world challenges. The trade-off between accuracy and interpretability is not a binary choice, but rather a spectrum, and the defeasible reasoning approach offers a valuable point along that spectrum, providing a balance between quantitative performance and qualitative understanding.

Future work should also focus on exploring hybrid approaches that combine the strengths of different techniques. For example, one could use knowledge graph embedding models to learn initial embeddings of geospatial entities and then use defeasible reasoning to refine these embeddings based on contextual information and domain knowledge. This could potentially lead to improved accuracy while still maintaining a degree of interpretability. Another promising direction is to explore the use of argumentation-based reasoning techniques to resolve conflicts between defeasible rules in a more nuanced and transparent way. Argumentation-based reasoning involves constructing arguments for and against different conclusions and then evaluating the strength of these arguments based on various criteria, such as the credibility of the sources, the logical validity of the reasoning, and the relevance of the evidence. This could provide a more flexible and adaptable approach to conflict resolution, allowing the system to take into account the specific context and characteristics of each situation.

authorship contribution statement

Bongjae Kwon: Writing – original draft, Resources, Investigation, Data curation, Methodology, Conceptualization, Formal analysis. **Kiyun Yu:** Writing – review & editing, Supervision, Funding acquisition, Project administration.

Informed consent

Informed consent was obtained from all individual participants included in the study

Ethics approval and consent to participate

This article does not contain any studies with human participants or animals performed by any of the authors.

Funding

This research was conducted with the support of the Ministry of Land, Infrastructure, and Transport (MOLIT) / Korea Agency for Infrastructure Technology Advancement (KAIA) (Project No. RS-2022-00143336, NTIS Project No. 1615013084).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The dataset used in this study is publicly available at: `https://github.com/bongjaekwon/geokgs-w-defeasible-reasoning`.

References

[1] Yan, B., Janowicz, K., Mai, G., & Gao, S. (2017). From ITDL to Place2Vec: Reasoning about place type similarity and relatedness by learning embeddings from augmented spatial contexts. *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL '17)*, Article 35, 1–10. https://doi.org/10.1145/3139958.3140054

[2] Koldasbayeva, D., Tregubova, P., Gasanov, M., et al. (2024). Challenges in data-driven geospatial modeling for environmental research and practice. *Nature Communications*, 15, 10700. https://doi.org/10.1038/s41467-024-55240-8

[3] Cao, K., Zhou, C., Church, R., Li, X., & Li, W. (2024). Revisiting spatial optimization in the era of geospatial big data and GeoAI. *International Journal of Applied Earth Observation and Geoinformation*, 129, 103832. https://doi.org/10.1016/j.jag.2024.103832

[4] Ihnaini, B., Abuhaija, B., Mills, E. A., & Mahmuddin, M. (2024). Semantic similarity on multimodal data: A comprehensive survey with applications. *Journal of King Saud University - Computer and Information Sciences*, 36(10), 102263. https://doi.org/10.1016/j.jksuci.2024.102263

[5] Mai, G., Hu, Y., Gao, S., Cai, L., Martins, B., Scholz, J., Gao, J., & Janowicz, K. (2022). Symbolic and subsymbolic GeoAI: Geospatial knowledge graphs and spatially explicit machine learning. *Transactions in GIS*, 26, 3118–3124. https://doi.org/10.1111/tgis.13012

[6] Lee, W. J., & Lauw, H. W. (2024). Latent representation learning for geospatial entities. *ACM Transactions on Spatial Algorithms and Systems*, 10(4), Article 32, 31 pages. https://doi.org/10.1145/3663474

[7] Moustafa, A. A. (1989). *Architectural representation and meaning: Towards a theory of interpretation* (Master's thesis). Massachusetts Institute of Technology, Department of Architecture. http://hdl.handle.net/1721.1/78999

[8] Mai, G., Huang, W., Sun, J., Song, S., Mishra, D., Liu, N., Gao, S., Liu, T., Cong, G., Hu, Y., Cundy, C., Li, Z., Zhu, R., & Lao, N. (2024). On the opportunities and challenges of foundation models for GeoAI (Vision Paper). *ACM Transactions on Spatial Algorithms and Systems*, 10(2), Article 11, 46 pages. https://doi.org/10.1145/3653070

[9] Delmelle, E. M., Desjardins, M. R., Jung, P., Owusu, C., Lan, Y., Hohl, A., & Dony, C. (2022). Uncertainty in geospatial health: Challenges and opportunities ahead. *Annals of Epidemiology*, 65, 15–30. https://doi.org/10.1016/j.annepidem.2021.10.002

[10] Lee, J.-G., & Kang, M. (2015). Geospatial big data: Challenges and opportunities. *Big Data Research*, 2(2), 74–81. https://doi.org/10.1016/j.bdr.2015.01.003

[11] Koons, R. (2022). Defeasible reasoning. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2022 Edition). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2022/entries/reasoning-defeasible/

[12] Jiang, B., Wan, G., Xu, J., Li, F., & Wen, H. (2018). Geographic knowledge graph building extracted from multi-sourced heterogeneous data. *Acta Geodaetica et Cartographica Sinica*, 47(8), 1051–1061. https://doi.org/10.11947/j.AGCS.2018.20180113

[13] Longo, L., Rizzo, L., & Dondio, P. (2021). Examining the modelling capabilities of defeasible argumentation and non-monotonic fuzzy reasoning. *Knowledge-Based Systems*, 211, 106514. https://doi.org/10.1016/j.knosys.2020.106514

[14] Besnard, P., Moinard, Y., Pereira, W., Clarke, M., & Wilson, N. (1993). DRUMS: Defeasible reasoning and uncertainty management systems. *AI Communications*, 6(1), 27–46.

[15] Longo, L., Brcic, M., Cabitza, F., Choi, J., Confalonieri, R., Del Ser, J., Guidotti, R., Hayashi, Y., Herrera, F., Holzinger, A., Jiang, R., Khosravi, H., Lecue, F., Malgieri, G., Páez, A., Samek, W., Schneider, J., Speith, T., & Stumpf, S. (2024). Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions. *Information Fusion*, 106, 102301. https://doi.org/10.1016/j.inffus.2024.102301

[16] Diogo, V., Bürgi, M., Debonne, N., Helfenstein, J., Levers, C., Swart, R., & Verburg, P. H. (2023). Geographic similarity analysis for Land System Science: Opportunities and tools to facilitate knowledge integration and transfer. *Journal of Land Use Science*, 18(1), 227–248. https://doi.org/10.1080/1747423X.2023.2218372

[17] Antonelli, A. G. (2004). Logic. In L. Floridi (Ed.), *The Blackwell Guide to the Philosophy of Computing and Information*. https://doi.org/10.1002/9780470757017.ch20

[18] Billington, D. (2008). Propositional clausal defeasible logic. In S. Hölldobler, C. Lutz, & H. Wansing (Eds.), *Logics in Artificial Intelligence. JELIA 2008* (Lecture Notes in Computer Science, vol. 5293). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-87803-2_5

[19] Andrews, R. (2003). An automated rule refinement system (PhD thesis). Queensland University of Technology. https://eprints.qut.edu.au/15788/

[20] Dash, T., Chitlangia, S., Ahuja, A., et al. (2022). A review of some techniques for inclusion of domain-knowledge into deep neural networks. *Scientific Reports*, 12, 1040. https://doi.org/10.1038/s41598-021-04590-0

[21] Li, Q., Li, Y., Gao, J., Zhao, B., Fan, W., & Han, J. (2014). Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation. In *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data (SIGMOD '14)* (pp. 1187–1198). Association for Computing Machinery. https://doi.org/10.1145/2588555.2610509

[22] Mai, G., Xie, Y., Jia, X., Lao, N., Rao, J., Zhu, Q., Liu, Z., Chiang, Y.-Y., & Jiao, J. (2025). Towards the next generation of Geospatial Artificial Intelligence. *International Journal of Applied Earth Observation and Geoinformation*, 136, 104368. https://doi.org/10.1016/j.jag.2025.104368

[23] Tian, L., Zhou, X., Wu, Y.-P., Zhou, W.-T., Zhang, J.-H., & Zhang, T.-S. (2022). Knowledge graph and knowledge reasoning: A systematic review. *Journal of Electronic Science and Technology*, 20(2), 100159. https://doi.org/10.1016/j.jnlest.2022.100159

[24] Lv, X., Gu, Y., Han, X., Hou, L., Li, J., & Liu, Z. (2019). Adapting meta knowledge graph information for multi-hop reasoning over few-shot relations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 3376–3381). Association for Computational Linguistics.

[25] Xian, Y., Fu, Z., Muthukrishnan, S., de Melo, G., & Zhang, Y. (2019). Reinforcement knowledge graph reasoning for explainable recommendation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'19)* (pp. 285–294). Association for Computing Machinery. https://doi.org/10.1145/3331184.3331203

[26] Chen, J., Deng, S., & Chen, H. (2017). CrowdGeoKG: Crowdsourced Geo-Knowledge Graph. In J. Li, M. Zhou, G. Qi, N. Lao, T. Ruan, & J. Du (Eds.), *Knowledge Graph and Semantic Computing. Language, Knowledge, and Intelligence. CCKS 2017* (Communications in Computer and Information Science, vol. 784). Springer, Singapore. https://doi.org/10.1007/978-981-10-7359-5_17

[27] Nute, D. (2003). Defeasible logic. In O. Bartenstein, U. Geske, M. Hannebauer, & O. Yoshie (Eds.), *Web Knowledge Management and Decision Support. INAP 2001* (Lecture Notes in Computer Science, vol. 2543). Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-36524-9_13

[28] Delrieux, C. (2004). Abductive inference in defeasible reasoning: A model for research programmes. *Journal of Applied Logic*, 2(4), 409–437. https://doi.org/10.1016/j.jal.2004.07.003

[29] Dai, X. L., Zhu, Y. Q., Yang, J., et al. (2022). Research and implementation of geospatial data similarity calculation method. *Journal of Global Change Data & Discovery*, 6(4), 501–512. https://doi.org/10.3974/geodp.2022.04.01

[30] Schwering, A. (2008). Approaches to semantic similarity measurement for geo-spatial data: A survey. *Transactions in GIS*, 12, 5–29. https://doi.org/10.1111/j.1467-9671.2008.01084.x

[31] Xiao, J., & He, Z. (2017). A novel approach to semantic similarity measurement based on a weighted concept lattice: Exemplifying geo-information. *ISPRS International Journal of Geo-Information*, 6(11), 348. https://doi.org/10.3390/ijgi6110348

[32] van Kreveld, M., Miltzow, T., Ophelders, T., Sonke, W., & Vermeulen, J. L. (2022). Between shapes, using the Hausdorff distance. *Computational Geometry*, 100, 101817. https://doi.org/10.1016/j.comgeo.2021.101817

[33] Rodriguez, M. A., & Egenhofer, M. J. (2003). Determining semantic similarity among entity classes from different ontologies. *IEEE Transactions on Knowledge and Data Engineering*, 15(2), 442–456. https://doi.org/10.1109/TKDE.2003.1185844

[34] Bordes, A., Usunier, N., Garcia-Durán, A., Weston, J., & Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'13)* (Vol. 2, pp. 2787–2795). Curran Associates Inc.

[35] Guo, S., Wang, Q., Wang, L., Wang, B., & Guo, L. (2016). Jointly embedding knowledge graphs and logical rules. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 192–202). Association for Computational Linguistics.

[36] Das, R., Dhuliawala, S., Zaheer, M., Vilnis, L., Durugkar, I., Krishnamurthy, A., Smola, A., & McCallum, A. (2018). Go for a walk and arrive at the answer: Reasoning over paths in knowledge bases using reinforcement learning. *arXiv:1711.05851* [cs.CL]. https://doi.org/10.48550/arXiv.1711.05851

[37] Bassiliades, N., Antoniou, G., & Governatori, G. (2007). Proof explanation in the DR-DEVICE system. In *Web Reasoning and Rule Systems. RR 2007. Lecture Notes in Computer Science, vol 4524* (pp. 273-286). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-72982-2_19

[38] Bryant, D., & Krause, P. (2008). A review of current defeasible reasoning implementations. *The Knowledge Engineering Review*, 23(3), 227-260. https://doi.org/10.1017/S0269888908001318

[39] Kravari, K., Papatheodorou, C., Antoniou, G., & Bassiliades, N. (2011). Reasoning and proofing services for semantic web agents. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011)* (pp. 2662-2667). https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-443

[40] Stoutenburg, S., Obrst, L., Nichols, D., Samuel, K., & Franklin, P. (2006). Applying semantic rules to achieve dynamic service-oriented architectures. In *Proceedings of the 2006 IEEE International Symposium on Rule Interchange and Application (RuleML 2006)* (pp. 75-82). https://doi.org/10.1109/RULEML.2006.4

[41] Powers, D., & Ailab, (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technology, 2*, 2229-3981. https://doi.org/10.9735/2229-3981

[42] Rizzo, L., & Longo, L. (2018). A qualitative investigation of the degree of explainability of defeasible argumentation and non-monotonic fuzzy reasoning. In *Proceedings of the 26th AIAI Irish Conference on Artificial Intelligence and Cognitive Science* (pp. 138-149). https://doi.org/10.21427/tby8-8z04

[43] Simeonw. (2025). Geo-Wikipedia-GeoNames Dataset. https://huggingface.co/datasets/simeonw/geo_wikipedia_geonames

[44] Cholewinski, P., Marek, V. W., & Truszczynski, M. (1996). Default reasoning system DeReS. In *Proceedings of the Fifth International Conference on Principles of Knowledge Representation and Reasoning (KR'96)* (pp. 518–528). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[45] Ali, M., Berrendorf, M., Hoyt, C. T., Vermue, L., Sharifzadeh, S., Tresp, V., & Lehmann, J. (2021). PyKEEN 1.0: A Python Library for Training and Evaluating Knowledge Graph Embeddings. *Journal of Machine Learning Research, 22*(82), 1–6. http://jmlr.org/papers/v22/20-825.html

[46] Anil, R., Borgeaud, S., Alayrac, J.-B., Yu, J., Soricut, R., Schalkwyk, J., Dai, A. M., Hauth, A., Millican, K., Silver, D., Johnson, M., *et al.* (2023). Gemini: A Family of Highly Capable Multimodal Models. *arXiv:2312.11805 [cs.CL]*. https://doi.org/10.48550/arXiv.2312.11805