Indonesian Journal of Electrical Engineering and Computer Science
Vol. 41, No. 3, March 2026, pp. 1040~1048
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v41.i3.pp1040-1048

Ontology-based semantic link prediction for enhancing academic collaboration through knowledge management

Pham Thi Thu Thuy[1], Thinh Thi Thuy[2]
[1]Faculty of Information Technology, Nha Trang University, Nha Trang, Vietnam
[2]Naval Academy, Nha Trang, Vietnam

Article Info

Article history:

Received Jun 29, 2025
Revised Jan 13, 2026
Accepted Feb 27, 2026

Keywords:

Academic collaboration
Ontology
Scholarly knowledge graph
Semantic integration
Semantic link prediction

ABSTRACT

This paper introduces a novel ontology-based semantic link prediction framework that unifies structural, temporal, and semantic signals from heterogeneous scholarly sources to enhance academic collaboration forecasting. By integrating AMiner, DBLP, and Mendeley datasets into a unified SKOS- and Dublin Core-aligned ontology, the framework enables semantic enrichment, cross-source reasoning, and contextualized link prediction. Unlike previous studies that focus solely on structural features or basic content similarity, our approach leverages ontology-based semantic feature engineering and graph-based learning for robust and interpretable predictions. Experimental results show that random forest and graph neural networks significantly outperform traditional models, achieving high accuracy and ranking precision. This work contributes to knowledge management by enabling expert recommendation, trend identification, and semantic integration for strategic academic planning.

Corresponding Author:

Pham Thi Thu Thuy
Faculty of Information Technology, Nha Trang University
02 Nguyen Dinh Chieu Street, Nha Trang, Vietnam
Email: thuthuy@ntu.edu.vn

1. INTRODUCTION

Collaboration among researchers plays a critical role in driving scientific innovation and generating new knowledge. As research communities expand and diversify, understanding the patterns and dynamics of co-authorship becomes increasingly important for institutions aiming to support effective knowledge sharing and strategic partnerships. Social networks constructed from co-authorship data offer a valuable foundation for exploring collaboration behaviors and identifying potential academic connections.

From a knowledge management perspective, the ability to predict future collaborations and uncover latent relationships within scholarly networks provides several organizational advantages. These include enhanced expert recommendation systems, early detection of emerging research clusters, and improved integration of heterogeneous knowledge sources. By leveraging semantic technologies and link prediction models, institutions can better manage intellectual capital, facilitate cross-disciplinary collaboration, and support evidence-based decision-making in research strategy and funding allocation.

Conventional link prediction techniques primarily rely on structural features such as network topology, node similarity, or historical co-authorship trends. While effective in some cases, these approaches often struggle in sparse or fragmented academic networks, where a deeper semantic context is required to uncover meaningful links. To address this limitation, ontologies offer a promising alternative by providing a formal, machine-readable representation of entities, relationships, and domain-specific knowledge. Through ontological reasoning, it becomes possible to enrich the representation of social networks with semantic metadata, enabling more accurate and interpretable link predictions.

However, existing approaches often lack the integration of multi-source scholarly metadata into a coherent ontological structure and rarely exploit both semantic enrichment and temporal collaboration history. In particular, most link prediction models either focus on structural features [1], [2] or use limited semantic signals such as keyword or topic similarity [3]–[6], and do not incorporate readership patterns, discipline taxonomies, or publication-type similarity within a unified framework.

To address these gaps, this study proposes an ontology-based semantic link prediction framework that integrates co-authorship data from three prominent academic sources - AMiner [7], DBLP [8], and Mendeley [9]. Our approach combines structural features such as historical collaboration with semantically enriched data on affiliations, research interests, publication types, readership patterns, and academic disciplines. The ontology is aligned with SKOS [10] and Dublin Core [11] standards to ensure interoperability and consistency across datasets. Community detection using the Louvain algorithm [12], SPARQL-based semantic feature extraction, and supervised machine learning models are used to evaluate predictive performance and demonstrate the value of this method for knowledge-driven decision support in academic settings.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 describes the proposed ontology-based method and feature construction. Section 4 presents the experimental setup, results, and comparative analysis. Section 5 discusses the implications and limitations of the findings. Section 6 concludes the paper and outlines directions for future research.

2. RELATED WORK

Link prediction in social and academic networks has been widely explored using structural, content-based, and semantic approaches. Traditional methods rely on network topology, such as common neighbors or graph-based similarity measures, often combined with supervised learning models [1], [2]. For instance, Hasan *et al.* [3] extracted structural and content-based features for supervised link prediction in academic networks, while Wohlfarth and Ichise [4] enhanced prediction using semantic keyword matching from paper titles. Similarly, Sachan and Ichise [5] introduced abstract keyword match count (AKMC) using Jaccard similarity, and Chuan *et al.* [6] proposed LDAcosin, a topic modeling approach for link prediction.

Several studies have also applied machine learning with domain knowledge. Hassan [13] developed a link prediction system based on research interests and affiliations using Python, R. Cho and Yu [14] explored interdisciplinary collaboration prediction using graph-based models. More recent studies focus on semantic link prediction in ontologies. Chen *et al.* [15], Wan *et al.* [16] applied contextual embeddings and ontology completion techniques. Ma *et al.* [17] proposed a deep learning model that integrates graph distance and entity semantics to infer subsumption links.

Ontology construction has also gained attention for domain-specific knowledge management. Researchers have applied ontologies to education [18], administration [19], forestry [20], and climate [21]. Others focus on ontology-enhanced information retrieval [22]–[24], especially in Vietnamese contexts [25]–[27]. In big data and multimedia, ontology-driven retrieval has been applied to image datasets [28], [29] and Wikipedia/DBPedia-derived ontologies [30].

In knowledge management, semantic technologies are increasingly applied for strategic reasoning and decision support. Mohammad *et al.* [31] proposed a knowledge graph-based system for enterprise knowledge services, reinforcing the utility of structured semantic models in organizational learning, an idea central to our work.

In summary, prior studies have effectively exploited structural features [1], [2], content and topic similarity [3]-[6], and ontology-based representations for retrieval and completion tasks [15]-[17], [22]-[30]. However, they generally do not integrate heterogeneous scholarly metadata (e.g., readership, discipline taxonomies, publication types) into a single SKOS - and Dublin Core–aligned ontology, and they rarely combine such semantic enrichment with temporal collaboration history for link prediction. Our work addresses this gap by constructing a unified academic ontology over AMiner, DBLP, and Mendeley and by empirically evaluating the added value of ontology-based semantic and temporal features against structural baselines.

3. METHOD

3.1. Graph construction and ontology integration

To build a comprehensive academic knowledge graph, we integrate three major datasets: AMiner, DBLP, and Mendeley. Each offers complementary metadata - AMiner provides co-authorship and citation data; DBLP contributes structured publication and venue information; and Mendeley adds readership, document types, and discipline tags.

These datasets are unified through a shared ontology developed in Protégé using the Cellfie plugin [32]. The ontology includes core classes such as Author, Publication, Affiliation, ResearchInterest, DocumentType, Venue, and AcademicDiscipline, as visualized in Figure 1. Semantic relationships among these entities are modeled using object properties like hasAffiliation, hasDiscipline, and hasPublication (Figure 2).

To ensure interoperability, we align key components with established standards: SKOS is used for controlled concepts (e.g., DocumentType, AcademicDiscipline), and Dublin Core for metadata annotations (e.g., dcterms: type, dcterms: publisher). This semantic alignment enables reasoning across heterogeneous data and supports integration with scholarly platforms like OpenAlex and ORCID.

The ontology structure supports the extraction of semantic features for link prediction and facilitates rule-based data import (Figure 3). For instance, the imported Author instance in Figure 3 shows linked affiliations, co-authors, and research interests, enriched with metrics like hasHIndex and hasCitationCount. This unified, standards-aligned ontology forms the foundation for semantic collaboration forecasting.

Before feature extraction, we used Protégé's built-in consistency checking to ensure that the ontology contained no obvious modeling conflicts or unsatisfied classes. This lightweight validation step helps ensure that the derived semantic features reflect a coherent conceptualization of the academic domain.

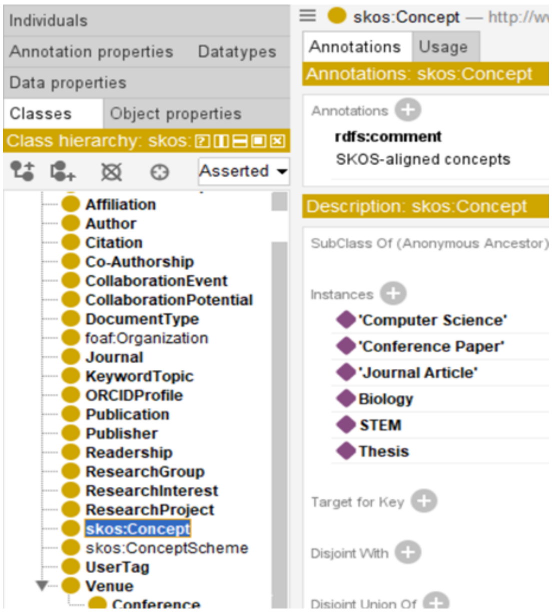

Figure 1. Class visualization of integrated ontology

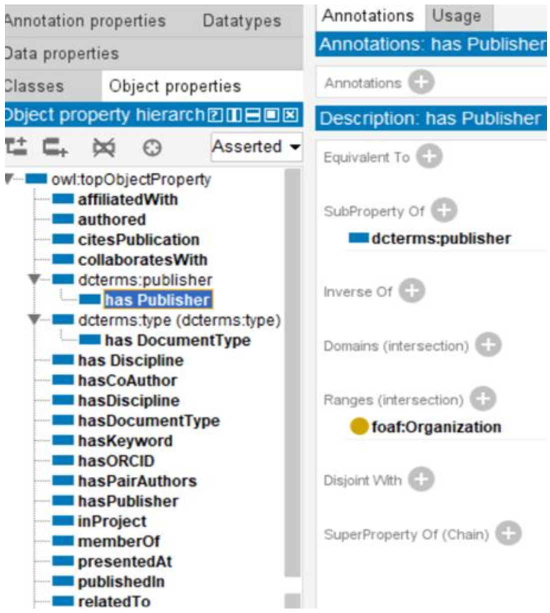

Figure 2. Object property visualization

3.2. Semantic feature engineering

From the enriched ontology, we derive a set of semantic and temporal features that capture collaboration dynamics and scholarly context. These include hasCommonAffiliation, which represents the degree of institutional overlap between author pairs; hasCommonInterest, a similarity score based on shared research interests or topics; hasReaderOverlap, derived from Mendeley data to reflect common readership patterns; hasPublicationTypeSimilarity, which assesses alignment in document types such as journal articles or conference papers; and hasDisciplineSimilarity, which is based on SKOS-aligned academic discipline tags. These similarities are encoded as data properties which are presented in the Figure 4.

Figure 4 shows the data property hierarchy of the integrated ontology, emphasizing key attributes linked to academic entities. The highlighted property, hasCommonAffiliation, stores a numerical value (*xsd:float*) representing institutional overlap between co-authors. Other properties such as hasCitationCount, hasHIndex, hasPublicationYear, and hasCommonInterest capture quantitative aspects of scholarly activity and author similarity. These structured features, derived from AMiner, DBLP, and Mendeley, support consistent extraction and improve the explainability of link prediction models.

Temporal features in the ontology include hasStatus0 to hasStatus3, which represent the number of co-authorships in the current year and the previous three years, respectively, capturing trends in collaboration over time. These features are extracted and numerically encoded from ontology instances of the CoAuthorship class and related entities.

After building the ontology structure from 3 datasets, data import is done by building a set of rules to transfer data from Excel file into ontology through Celfie tool [32]. The data import result is shown in Figure 3. Figure 3 presents a snapshot of individual data imported into the integrated ontology, specifically showcasing an instance of the Author class. This instance (Author ID: 1000228) is semantically enriched with both object property assertions, such as hasCoAuthor, authored, affiliatedWith, and relatedTo, and data property assertions, including hasName, hasInterestArea, hasCitationCount, hasHIndex, and hasPaperCount. The object properties establish relationships between this author and other entities like publications, affiliations, and co-authors, while data properties capture descriptive and metric details such as research interests (e.g., “digital filter”, “expert control system”) and publication impact indicators (e.g., hasHIndex = 1, hasCitationCount = 2). This structured representation supports advanced semantic reasoning and feature extraction for link prediction. The successful integration of instance data from sources like AMiner and DBLP demonstrates the ontology’s capacity to unify heterogeneous scholarly metadata into a consistent, queryable knowledge graph.

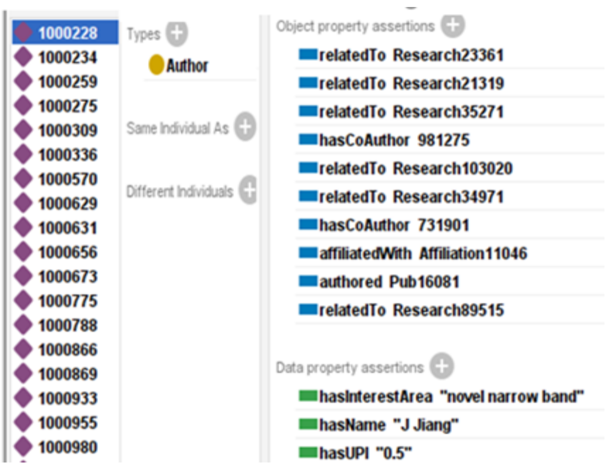

Figure 3. Data import result of the integrated ontology

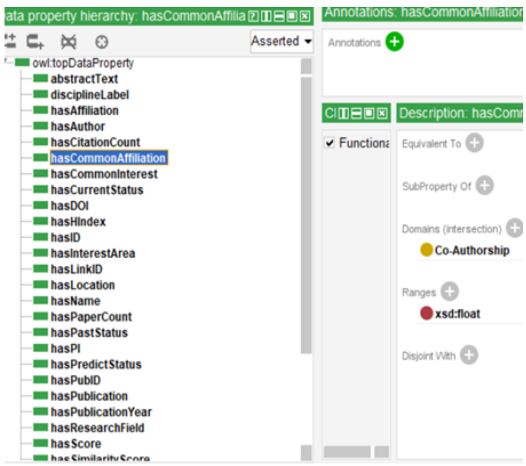

Figure 4. Data property visualization

4. RESULTS AND DISCUSSION

This section presents the experimental setup, evaluation metrics, and performance results of our ontology-based link prediction framework. The evaluation is designed to assess the effectiveness of integrating heterogeneous scholarly datasets and semantic features in enhancing predictive accuracy and knowledge discovery.

4.1. Feature encoding and learning setup

Each author pair was represented by a feature vector that combines both structural and semantic dimensions to capture the complexity of academic collaboration. The temporal features include hasStatus3, hasStatus2, and hasStatus1, which reflect the number of co-authorships over the past three years. The semantic features consist of hasCommonAffiliation, hasCommonInterest, hasDisciplineSimilarity, hasReaderOverlap, and hasPublicationTypeSimilarity, all of which are derived from the enriched ontology to represent institutional, topical, disciplinary, readership, and publication-type similarities between authors.

These features were automatically extracted from the ontology using SPARQL queries and converted into numerical feature vectors for input to machine learning models. Figure 5 provides an example of a SPARQL query used to compute common affiliation scores. Models were trained using scikit-learn for classical classifiers, and graph neural network variants were implemented using a standard deep learning library for graph-based reasoning. The final dataset includes 3312 authors and 14268 author pairs, with 7134 positive (existing collaboration) and 7134 negative (non-collaboration) instances. We randomly split the data into 70% training, 15% validation, and 15% test sets, stratified by class label to preserve the positive/negative ratio. For traditional machine learning models, we used scikit-learn implementations and tuned key hyperparameters (such as regularization strength for logistic regression and number of trees and maximum depth for random forest) on the validation set. For the GNN, we trained a two-layer architecture with non-linear activation and early stopping based on validation loss.

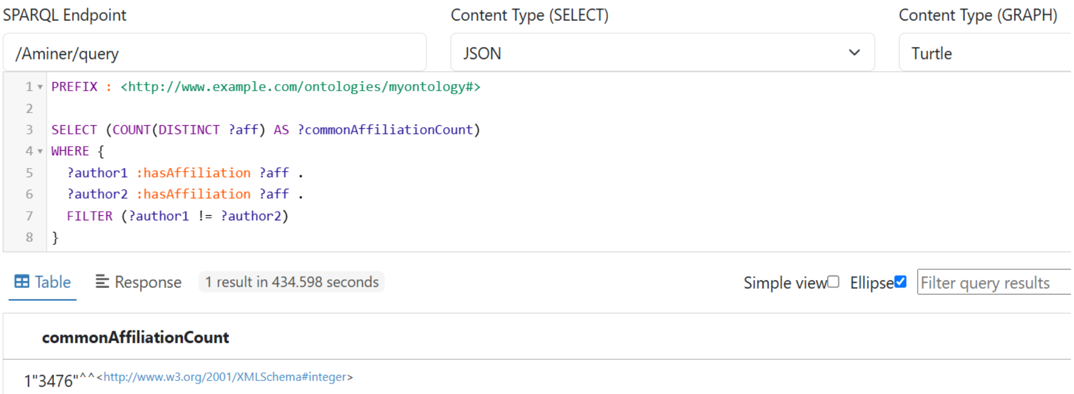

Figure 5. SPARQL to extract common affiliation

Figure 5 illustrates the execution of a SPARQL query designed to compute the number of common institutional affiliations between pairs of authors within the integrated ontology. The query selects all distinct affiliations (?aff) that are shared by any two authors (?author1 and ?author2) and applies a filter to ensure that only non-identical author pairs are considered. The query was executed on a SPARQL endpoint supporting the ontology at *http://www.example.com/ontologies/myontology#*. As shown in the result panel, the system identified 3,476 shared affiliation instances, indicating a high level of institutional overlap among the authors in the dataset. This feature, labeled commonAffiliationCount, is used as a semantic similarity metric in the link prediction model to enhance its ability to forecast future academic collaborations.

In practice, this query is part of a family of parameterized SPARQL templates that are iteratively executed over all candidate author pairs to compute semantic features such as common affiliation, shared interests, discipline similarity, and reader overlap. This design allows the feature extraction process to scale to large numbers of pairs while keeping the SPARQL logic modular and reusable.

4.2. Results and comparative analysis

Experiments demonstrate that incorporating ontology-based features significantly improves link prediction performance. The experimental results are presented in Figures 6-9, which summarize classification accuracy, F1-score, AUC, and ranking metrics across the evaluated models. Together, these figures highlight the impact of both semantic and temporal features on collaboration forecasting.

Figure 6 shows model performance with a 4-feature setup, including current co-authorship and basic semantic attributes like affiliation and research interest. Logistic regression and random forest yield strong results (84.3% and 86.2% accuracy), confirming the effectiveness of even minimal semantic features.

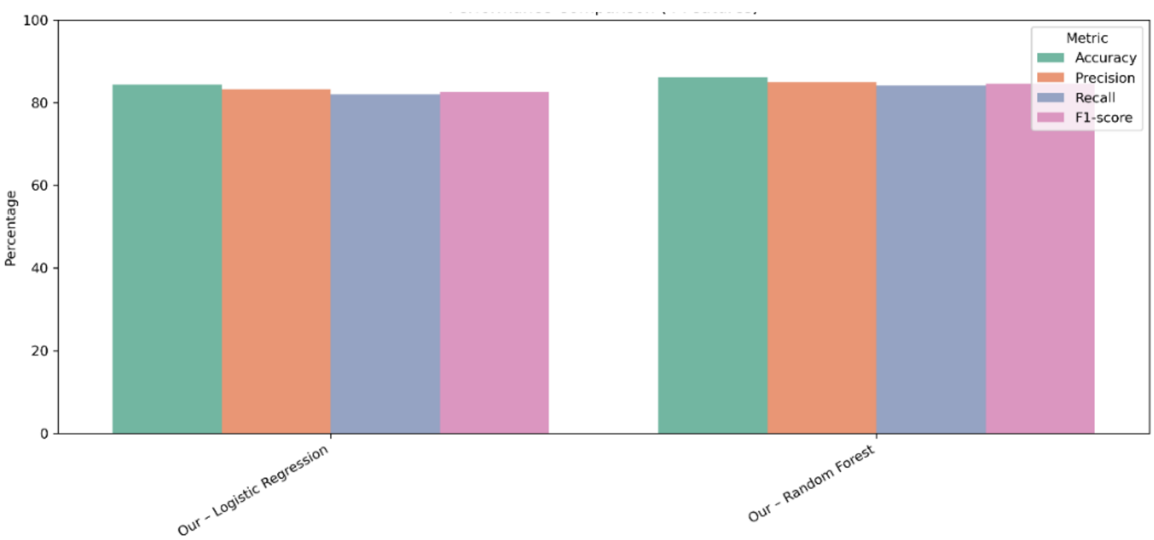

Figure 6. Performance comparison of models using the 4-feature configuration

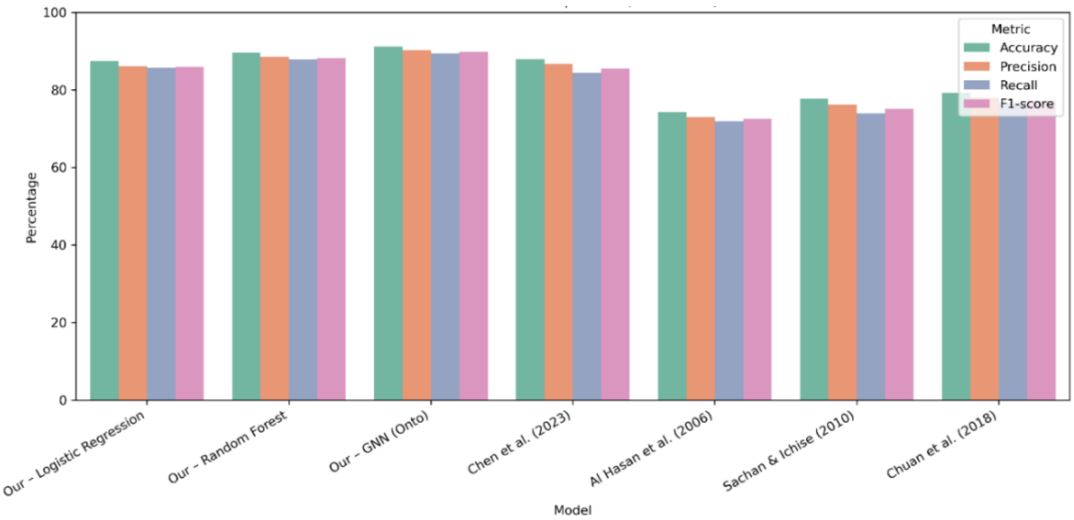

Figure 7. Performance comparison of models using the 5-feature configuration

Figure 7 compares models using a 5-feature configuration that includes three years of historical collaboration. Adding temporal features boosts all metrics, with GNN and random forest outperforming baselines-GNN achieves 91.2% accuracy and 89.8% F1-score. The improvement over traditional models for co-authorship link prediction, such as the supervised structural approach of Al Hasan *et al.* [3], the semantic enhancements of Sachan and Ichise [11], and the hybrid content-based method of Chuan *et al.* [6], highlights the advantage of combining temporal and ontology-based features. The overall performance is also consistent with recent semantic embedding - based approaches such as Chen *et al.* [15], indicating that our framework is competitive within the current state of the art in semantic link prediction for academic networks.

Figure 8 presents AUC scores, clearly favoring ontology-enhanced models (AUC 0.92-0.96) over structure-only baselines (<0.86). This underscores the benefit of incorporating semantic features for better class separability and prediction reliability.

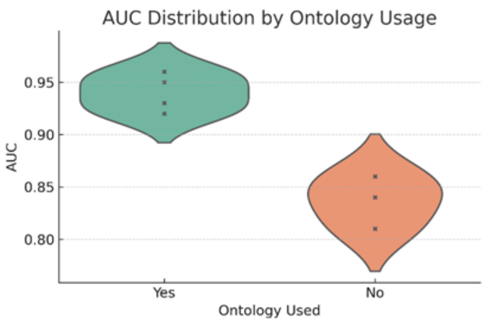

Figure 8. AUC distribution by ontology usage

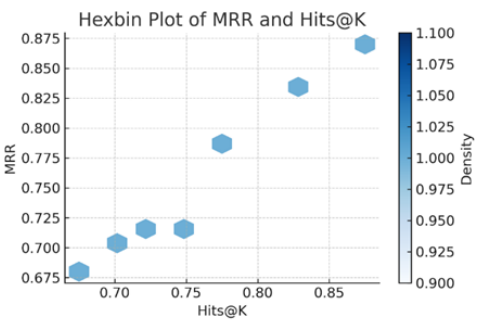

Figure 9. MRR and Hits@K values across models

Figure 9 displays a hexbin plot showing the relationship between mean reciprocal rank (MRR) and Hits@K for all evaluated models. Darker areas indicate higher prediction density, revealing a clear upward trend-models with higher Hits@K also tend to achieve higher MRR, reflecting strong ranking precision. Ontology-based models, especially those using semantic and temporal features, cluster in the upper-right region, highlighting their effectiveness in identifying not just likely but highly relevant collaborations.

Experiments confirm that five-feature models, which include historical co-authorship (hasStatus3 to hasStatus1), significantly enhance prediction accuracy. Random forest and logistic regression consistently outperform simpler methods, while GNNs achieve the highest performance, showing promise for future work.

Across all metrics, ontology-enhanced models outperform traditional approaches. Random forest and logistic regression, when enriched with semantic and temporal features, approach 90% accuracy and yield strong MRR and AUC scores. In contrast, models based solely on structure or content similarity perform notably worse. These findings underscore the value of combining ontological reasoning with collaboration history to improve link prediction, particularly in sparse academic networks.

Compared with ontology-based and semantic link prediction methods in the literature [3]-[6], [15], [17], our best models achieve competitive or higher accuracy and ranking performance while operating on a richer feature space that jointly encodes structure, time, and SKOS-aligned semantics. Although exact experimental settings differ across studies, the consistent gains of ontology-enhanced features over our own structural baselines suggest that semantic modeling adds complementary information beyond what is captured by network metrics alone.

4.3. Discussion

The experimental results show that integrating semantic and temporal features significantly improves the prediction of academic collaborations compared to structure-only baselines. Ontology-enhanced models demonstrate higher accuracy, F1-score, AUC, and ranking metrics, confirming that semantic enrichment provides meaningful context not captured by traditional topological measures. These findings are consistent with earlier work that incorporates content and semantic information into link prediction [3]-[6], [15], but our study further combines history and SKOS-aligned disciplinary knowledge derived from a unified ontology.

From an application perspective, the framework is relevant for institutions that aim to build data-driven researcher recommendation systems and to monitor the evolution of research communities. Semantic features such as affiliation, discipline similarity, and readership overlap help identify promising collaboration opportunities that are not obvious from structure alone, thereby supporting expert recommendation, team formation, and the detection of emerging research clusters. These capabilities align with the broader role of semantic technologies in knowledge management and decision support [31].

This work has several limitations that suggest directions for future research. First, the ontology is not yet integrated with explicit temporal ontologies such as OWL-Time, which would enable more expressive reasoning about the evolution of collaborations. Second, we did not exhaustively explore alternative graph neural architectures or large language model-based representations, which could further improve performance. Third, real-world evaluation in institutional settings (e.g., through user studies or deployment in research support tools) remains an important open step.

In summary, the results indicate that combining multi-source semantic integration with machine learning provides an effective and explainable framework for academic collaboration prediction, while also opening opportunities for richer reasoning and validation in future work.

5. CONCLUSION

This study presents an ontology-based semantic link prediction framework that addresses the growing need for intelligent collaboration forecasting in academic environments. By integrating co-authorship data from AMiner, DBLP, and Mendeley into a unified ontology, and aligning it with SKOS and Dublin Core standards, we constructed a rich semantic representation of scholarly relationships and research contexts. The framework combines structural, temporal, and semantic features to infer potential collaborations, leveraging both rule-based reasoning and machine learning techniques.

The results demonstrate that our approach significantly improves link prediction performance, particularly in sparse networks where traditional structural methods are less effective. Beyond predictive accuracy, the framework contributes meaningfully to the field of knowledge management. It enables enhanced expert recommendation by capturing multi-dimensional similarity across authors, supports the early identification of emerging research areas through semantic trend analysis, and facilitates the integration of heterogeneous academic data for organizational learning.

By bridging social network analysis, ontology engineering, and machine learning, this research introduces an ontology-centered framework for academic link prediction that integrates multi-source academic data into a structured, standards-aligned representation and combines semantic enrichment with graph-based learning. Compared with traditional models that rely solely on network topology or basic content similarity, the proposed approach offers improved predictive performance and a more interpretable view of collaboration patterns. These contributions support the development of intelligent research support tools and provide a foundation that can be extended with more advanced reasoning and representation techniques in future work.

FUNDING INFORMATION

This research is funded by Nha Trang University for science and technology under grant number TR2025-13-05.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Pham Thi Thu Thuy	✓	✓	✓		✓			✓	✓	✓		✓	✓	
Thinh Thi Thuy						✓	✓			✓	✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL

This research did not involve any experiments on humans or animals. Therefore, ethical approval was not required. The study was conducted in full compliance with institutional and national research policies and guidelines. All data used were obtained from publicly available academic sources (AMiner, DBLP, and Mendeley), and no personal or sensitive information was collected or disclosed during the research process.

DATA AVAILABILITY

The data that support the findings of this study are publicly available from the following sources: AMiner: https://www.aminer.cn, DBLP: https://dblp.org, and Mendeley Dataset for Research Collaboration Analysis: https://doi.org/10.17632/z55c36y2nw.1.

The integrated ontology and derived feature datasets used for model training and evaluation are available from the corresponding author, P.T.T.T., upon reasonable request.

REFERENCES

[1] F. Gao, K. Musial, C. Cooper, and S. Tsoka, "Link prediction methods and their accuracy for different social networks and network metrics," *Scientific Programming*, vol. 2015, pp. 1–13, 2015, doi: 10.1155/2015/172879.

[2] W. Cukierski, B. Hamner, and B. Yang, "Graph-based features for supervised link prediction," in *The 2011 International Joint Conference on Neural Networks*, IEEE, Jul. 2011, pp. 1237–1244. doi: 10.1109/IJCNN.2011.6033365.

[3] M. Hasan *et al.*, "Link prediction using supervised learning link prediction using supervised learning *," *SDM06: workshop on link analysis, counter-terrorism and security*, 2006, [Online]. Available: https://www.researchgate.net/publication/277289291.

[4] T. Wohlfarth and R. Ichise, "Semantic and event-based approach for link prediction," in *International Conference on Practical Aspects of Knowledge Management*, 2008, pp. 50–61. doi: 10.1007/978-3-540-89447-6_7.

[5] M. Sachan and R. Ichise, "Using semantic information to improve link prediction results in network datasets," *International Journal of Engineering and Technology*, vol. 2, no. 4, pp. 334–339, 2010, doi: 10.7763/IJET.2010.V2.143.

[6] P. M. Chuan, L. H. Son, M. Ali, T. D. Khang, L. T. Huong, and N. Dey, "Link prediction in co-authorship networks based on hybrid content similarity metric," *Applied Intelligence*, vol. 48, no. 8, pp. 2470–2486, 2018, doi: 10.1007/s10489-017-1086-x.

[7] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su, "ArnetMiner: extraction and mining of academic social networks," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA: ACM, Aug. 2008, pp. 990–998, doi: 10.1145/1401890.1402008.

[8] DBLP Team. (n.d.), "DBLP computer science bibliography." Accessed: Mar. 29, 2025. [Online]. Available: https://dblp.org.

[9] M. Grčar, D. Mladenić, B. Fortuna, and M. Grobelnik, "Mendeley dataset for research collaboration analysis (Version 1) [Data set]," *Mendeley Data*. 2017. doi: 10.17632/z55c36y2nw.1.

[10] A. Miles and S. Bechhofer, "SKOS simple knowledge organization system reference," W3C Recommendation. Accessed: Mar. 29, 2025. [Online]. Available: https://www.w3.org/TR/skos-reference/.

[11] "Dublin Core™ Metadata Element Set, Version 1.1: Reference Description," Dublin core metadata initiative. Accessed: Mar. 29, 2025. [Online]. Available: https://www.dublincore.org/specifications/dublin-core/dces/.

[12] V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P10008, Oct. 2008, doi: 10.1088/1742-5468/2008/10/P10008.

[13] D. Hassan, "Supervised link prediction in co-authorship networks based on research performance and similarity of research interests and affiliations," in *International Conference on Machine Learning and Cybernetics*, IEEE, Jul. 2019, pp. 1–6. doi: 10.1109/ICMLC48188.2019.8949320.

[14] H. Cho and Y. Yu, "Link prediction for interdisciplinary collaboration via co-authorship network," *Social Network Analysis and Mining*, vol. 8, no. 1, p. 25, Dec. 2018, doi: 10.1007/s13278-018-0501-6.
[15] J. Chen, Y. He, Y. Geng, E. Jiménez-Ruiz, H. Dong, and I. Horrocks, "Contextual semantic embeddings for ontology subsumption prediction," *World Wide Web*, vol. 26, no. 5, pp. 2569–2591, Sep. 2023, doi: 10.1007/s11280-023-01169-9.
[16] Y. Wan, Z. Chen, Y. Liu, M. Packianather, and R. Wang, "Constructing industrial knowledge graph through ontology and link prediction," in *IEEE International Conference on Automation Science and Engineering*, IEEE, Aug. 2023, pp. 1–6, doi: 10.1109/CASE56687.2023.10260566.
[17] J. Ma *et al.*, "ELPKG: a high-accuracy link prediction approach for knowledge graph completion," *Symmetry*, vol. 11, no. 9, p. 1096, Sep. 2019, doi: 10.3390/sym11091096.
[18] G. Ashour, A. Al-Dubai, I. Romdhani, and D. Alghazzawi, "Ontology-based linked data to support decision-making within Universities," *Mathematics*, vol. 10, no. 17, 3148, 2022.
[19] Y. Bassil and P. Semaan, "Semantic-sensitive web information retrieval model for HTML documents," *International Journal of Computer Science Issues*, 2012.
[20] T. X. Truong, "Building an ontology to serve semantic search of forest animals in Da Nang city," Master's thesis, University of Da Nang, Da Nang, Vietnam, 2016.
[21] J. X. Shen, "Ontology-based semantic retrieval for management information system," *Applied Mechanics and Materials*, vol. 278–280, pp. 2069–2072, 2013.
[22] Y. Ma, "An ontology-based semantic retrieval model for information retrieval," *Journal of Information Science*, vol. 36, no. 6, pp. 707–720, 2010.
[23] N. T. Tuấn and H. T. T. Hà, "Building ontology for optional data query system," *Journal of Science and Technology*, vol. 1, pp. 128–132, 2017.
[24] T. T. T. Do and D. T. Nguyen, "A framework for Vietnamese text document retrieval system based on phrasal semantic analysis," *International Journal of Simulation: Systems, Science & Technology*, vol. 15, no. 4, pp. 61–66, 2014, doi: 10.5013/IJSSST.a.15.04.08.
[25] D. Yuvaraj, S. S. Alnuaimi, B. H. Rasheed, M. Sivaram, and V. Porkodi, "Ontology based semantic enrichment for improved information retrieval model," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 15s, pp. 70–77, 2024.
[26] T. T.-T. Do and D. T. Nguyen, "Phrasal semantic distance for Vietnamese textual document retrieval," *Journal of Computer Science and Cybernetics*, vol. 31, no. 3, pp. 185–199, 2015.
[27] V. T. Pham, H. D. Nguyen, T. Le, B. Nguyen, and Q. H. Ngo, "Ontology-based solution for building an intelligent searching system on traffic law documents," *arXiv preprint arXiv:2301.11252*, 2023.
[28] Y. Liu, D. Zhang, G. Lu, and W. Y. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recognition*, vol. 40, no. 1, pp. 262–282, 2007.
[29] U. Manzoor, M. A. Balubaid, B. Zafar, H. Umar, and M. S. Khan, "Semantic image retrieval: an ontology based approach," *International Journal of Advanced Research in Artificial Intelligence*, vol. 4, no. 4, pp. 1–8, 2015.
[30] T. Kawakami, T. Morita, and T. Yamaguchi, "Building Wikipedia ontology with more semi-structured information resources," in Proc. Joint International Semantic Technology Conference (JIST), 2017.
[31] Y. Mohammad, M. Nachouki, and E. A. Mohamed, "Knowledge management systems in business management using knowledge graphs and semantic technologies," *International Journal of Knowledge Management*, vol. 21, no. 1, pp. 1–29, Feb. 2025, doi: 10.4018/IJKM.369121.
[32] Protégé Project, "Cellfie plugin for Protégé," GitHub repository. [Online]. Available: https://github.com/protegeproject/cellfie-plugin.

BIOGRAPHIES OF AUTHORS

Pham Thi Thu Thuy received her B.Eng. degree in Computer Technology from Hanoi University of Science and Technology, Vietnam, in 2001. She obtained her M.Sc. and Ph.D. degrees in Computer Engineering from Kyung Hee University, South Korea, in 2006 and 2012, respectively. She is currently a senior lecturer at the Faculty of Information Technology, Nha Trang University, Vietnam. Her research interests include knowledge-based systems, ontology engineering, big data analytics, and intelligent learning support systems. Dr. Thuy is the author of several academic books, including XML Technology and Applications (2018) and Database Management System (2019), published by Scientific and Technical Publisher, Vietnam. Her research has been published in journals such as knowledge-based systems, wireless personal communications, and SN computer science. She can be contacted at email: thuthuy@ntu.edu.vn.

Thinh Thi Thuy received the B.Sc. degree in computer science from the Military Technical Academy, Vietnam in 2008. She is currently a master student in Information Technology at Nha Trang University. She is currently a lecturer at the Naval Academy, Vietnam. Her main research interests include ontology, machine learning, and big data processing. She can be contacted at email: thuy8384@gmail.com.