Completing Cultural Heritage Knowledge Graphs Using GraIL-Based Inductive Reasoning

Jaeyun Han*
Department of Mathematics
Konkuk University
Seoul, South Korea
dsa8978@konkuk.ac.kr

Jae Seung Kim*
Department of Mathematics
Konkuk University
Seoul, South Korea
dgd05037@konkuk.ac.kr

Jiwon Lee
Hyper-reality Metaverse Research Laboratory
Electronics and Telecommunication Research Institute
Daejeon, South Korea
ez1005@etri.re.kr

Hwijae Son
Department of Mathematics
Konkuk University
Seoul, South Korea
hwijaeson@konkuk.ac.kr

***Abstract*--Cultural heritage ontologies often lack critical metadata such as authorship, which limits their interpretability and practical utility. In this study, we frame authorship prediction as a link prediction problem and propose an inductive reasoning approach using Graph-based Inductive Learning (GraIL). Our method extracts local enclosing subgraphs around entity pairs and employs graph neural networks to learn structural patterns, enabling generalization to previously unseen artifacts and authors. Applied to a real-world cultural heritage dataset, the proposed approach achieves strong predictive performance, highlighting its effectiveness in enriching incomplete ontological data.**

***Keywords*—Knowledge graph, Link Prediction, Graph based Inductive Learning**

I. Introduction

Cultural heritage datasets frequently lack critical metadata, such as authorship, production dates, or provenance, which undermines their interpretability, reusability, and integration into downstream applications. These incomplete records constrain curatorial insight and pose significant challenges for researchers and practitioners seeking to analyze, classify, or recommend cultural artifacts at scale.

To address this limitation, we frame the task of predicting missing authorship as a link prediction problem, leveraging the structural relationships encoded in cultural heritage knowledge graphs. These graphs provide a natural representation of artifacts and their associated entities—such as authors, materials, styles, and periods—making them well-suited for relational inference. Specifically, we employ Graph-based Inductive Learning (GraIL) [2], which performs link prediction by reasoning over local subgraph structures. This inductive framework enables robust generalization to previously unseen artifacts and authors, effectively addressing cold-start challenges while maintaining scalability and accuracy in real-world cultural heritage scenarios.

II. Meothdology

A. Prooblem Fomulation

We formalize the authorship prediction task as a link prediction problem within the framework of a cultural heritage knowledge graph [1]. Let $G = (V, E)$ denote a graph, where V is the set of entities representing artifacts and authors, and $E \subseteq V \times V$ is the set of directed edges corresponding to authorship relations. The objective is to infer the existence of a missing edge between a given artifact and a candidate author, leveraging both the local and global structural properties of the graph.

B. Graph-based Inductive Learning (GraIL)

GraIL is an inductive link prediction framework designed to generalize to unseen entities. Instead of relying on static embeddings [3], GraIL leverages local enclosing subgraphs around a candidate link to perform reasoning. It follows a three-step process:

1) Extracting the k-hop enclosing subgraph between the source and target nodes
2) Applying Double-Radius Node Labeling (DRNL) [4] to encode node positions relative to both endpoints.
3) Using a graph neural network to compute a score for the link.

This inductive approach enables GraIL to effectively handle the cold-start problem, making it particularly suitable for sparse and evolving knowledge graphs such as those in the cultural heritage domain.

C. Methodology

We applied GraIL to a cultural heritage knowledge graph, where each node represents an entity such as an artifact, author, material, pattern, or category. The graph captures rich semantic relations between these entities, including authorship (artifact–author), composition (artifact–material), and style (artifact–pattern).

The primary task is to predict missing authorship links, i.e., to determine whether a given artifact was created by a specific author. To prepare training data, we sampled positive triplets of the form (artifact, created by, author) and generated negative samples by replacing either the artifact or the author. For each candidate link, we extracted a k-hop enclosing subgraph centered on the source and

*These authors contributed equally to this work.

target entities and applied DRNL to capture the structural role of each node within the subgraph.

The labeled subgraphs were encoded using a graph neural network, and the model was trained with a noise-contrastive hinge loss. We split the data into training, validation, and test sets, ensuring that some authors and artifacts in the test set did not appear during training to assess GraIL's inductive generalization capability. This setup allows the model to infer missing authorship relations based on contextual graph structure and to generalize to unseen entities—addressing the cold-start problem that commonly arises in cultural heritage datasets with incomplete metadata.

III. Experiments and Results

A. Experimental Setup

We evaluate our approach on a cultural heritage knowledge graph constructed from real-world datasets containing artifacts, authors, and their relationships. We split the dataset into 80% training, 10% validation, and 10% testing sets.

For local subgraph extraction, we use K = 3 to balance contextual richness and computational cost. The GraIL model employs a three-layer Graph Neural Network (GNN) [5] with hidden dimensions of 32. We train the model for 100 epochs with a batch size of 32, using the Adam optimizer and a learning rate of 0.01. The margin γ for the hinge loss is set to 10.

B. Evaluation Results

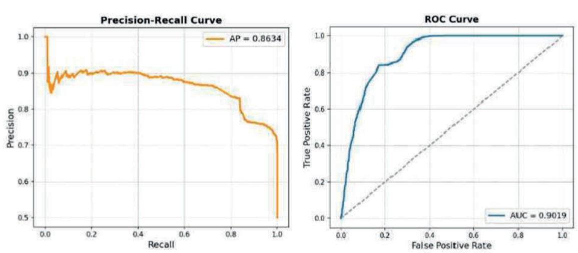

Fig. 1. Precision–recall and ROC curves for authorship link prediction.

We evaluate the performance of our approach using both threshold-based and ranking-based metrics. Precision-Recall and ROC curves are shown in Fig. 1, respectively. The area under the precision-recall curve (AUPRC) is 0.8634, indicating high precision across different recall levels. Similarly, the ROC curve shows strong separability, with an AUC of 0.9019.

TABLE 1. Ranking-based evaluation results for authorship link prediction.

MRR	Hits@1	Hits@5	Hits@10
0.5292	0.4313	0.6182	0.7895

Hits@K measures whether the correct answer appears among the top K predictions and is widely used in recommendation and link prediction tasks (see Table 1). The model achieves an MRR of 0.5292, indicating its ability to rank correct authorship links highly. Notably, Hits@10 reaches 0.7895, demonstrating strong top-K prediction performance. These results confirm that our GraIL-based method is effective at inferring missing authorship links in the cultural heritage knowledge graph.

The inductive nature of GraIL enables it to generalize to unseen authors and artifacts, effectively addressing cold-start challenges that frequently arise in real-world heritage datasets. Unlike transductive models that rely on fixed entity embeddings learned during training, GraIL reasons over the structural patterns of local subgraphs, allowing it to infer links involving entirely new nodes without retraining. This makes it particularly suitable for cultural heritage domains, where newly discovered artifacts or incomplete metadata are common and retraining is often impractical.

IV. Conclusion

In this study, we presented a novel approach for completing cultural heritage knowledge graphs by applying GraIL-based inductive reasoning for authorship prediction. By framing authorship prediction as a link prediction task and leveraging local enclosing subgraphs, our method enables effective generalization to previously unseen artifacts and authors, addressing the prevalent cold-start issue in cultural heritage datasets.

The success of our GraIL-based model opens avenues for future research. Potential directions include extending the framework to other metadata relationships (e.g., location, material, stylistic features) and incorporating multimodal data sources such as images or historical texts. Overall, this work contributes to the development of scalable, accurate, and generalizable systems for cultural heritage data completion and interpretation.

Acknowledgment

This research was supported by Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency(KOCCA) grant funded by the Ministry of Culture, Sports and Tourism(MCST) in 2023(Project Name: Development of storytelling AI technology for cultural heritage tailored to the various interests of users, Project Number: RS-2023-00220195, Contribution Rate: 100%)

References

[1] Kim, J., Lee, J., and Son, H., "Cultural Heritage Knowledge Graph and Recommender System," Proc. of the 15th International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea, 2024, DOI: 10.1109/ICTC62082.2024.10827123.

[2] Teru, K., Denis, E., and Hamilton, W. L., "Inductive Relation Prediction by Subgraph Reasoning," Proc. of the 37th International Conference on Machine Learning (ICML), 2020, pp. 9448–9457.

[3] Wang, Quan, Mao, Z., Wang, B., and Guo, L., "Knowledge Graph Embedding: A Survey of Approaches and Applications," IEEE Transactions on Knowledge and Data Engineering, Vol. 29, No. 12, 2017, pp. 2724–2743.

[4] M. Zhang and Y. Chen, "Link Prediction Based on Graph Neural Networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 31, 2018.

[5] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in Proceedings of the International Conference on Learning Representations (ICLR), A. Courville, Y. Bengio, and I. Goodfellow, Eds. Toulon, France, 2017. [Online]. Available: https://arxiv.org/abs/1609.02907