Towards Unsupervised Entity Alignment for Highly Heterogeneous Knowledge Graphs Technical Report

***Abstract*—This technical report contains the full set of experimental setting of the paper "Towards Unsupervised Entity Alignment for Highly Heterogeneous Knowledge Graphs".**

***Index Terms*—highly heterogeneous entity alignment, unsupervised, experiment setting**

I. Experiments

In this section, we first introduce the experimental setting[1] in Section I-A.

A. Experimental Setting

Datasets. We evaluate our approach using five entity alignment datasets, summarized in Table I.

Among these, `DBP15K(EN-FR)` and `DBP-WIKI` [3] are established benchmarks in the field of entity alignment. `DBP15K(EN-FR)` specifically focuses on cross-lingual entity alignment, while `DBP-WIKI` represents a larger-scale dataset for heterogeneous knowledge graphs. However, these KGs share similar structural properties, including density and structure similarity, with 100% overlapping ratios and comparable scales in terms of entities, relations, and facts.

`ICEWS-WIKI` and `ICEWS-YAGO` [1] exemplify highly heterogeneous entity alignment datasets[2], characterized by significant differences between their KGs. These differences are evident not only in the counts of entities, relations, and facts but also in structural characteristics (e.g., density and structural similarity) and their low overlapping ratio. These differences are evident not only in the counts of entities, relations, and facts but also in structural characteristics (e.g., density and structure similarity) and their low overlapping ratio. Notably, the number of anchors is not equivalent to the entity count, making the alignment task more challenging on these complex HHEA datasets.

`BETA` [4] is designed for new more complex temporal EA scenarios, comprising six realistic complex cases that more closely reflect real-world situations, multi-granularity temporal features, and etc. We will conduct further experiments on these cases in Section IV-F.

Baselines. Currently, no specific solutions exist for unsupervised HHEA. To establish a comprehensive baseline, we introduced 23 SOTA and classic baseline methods for extensive comparison:

- **Supervised Methods:** These methods require the use of 100% of the EA training set (we followed the 3:7 splitting ratio in training/ testing data). They are further categorized into: (1) translation-based nethods ("Trans." in Table II, such as MTransE [5], AlignE [6] and BootEA [6].). (2) GNN-based methods ("GNN" in Table II, such as GCN-Align [7], RDGCN [8], Dual-AMN [9], TEA-GNN [10], TREA [11], STEA [12] and Dual-Match [13].). (3) Other EA methods ("Other" in Table II, such as BERT [14], FuAlign [15], BERT-INT [16], PARIS [17], [18], Simple-HHEA [1], and ChatEA [2]).
- **Unsupervised & Self-Supervised Methods:** These methods do not use any EA training set data (e.g., MultiKE [19], SelfKG [20] and ZeroEA [21]). "*" represents that the model leverages the top 1 pseudo-labeling provided by `Candidate Entity Retrieval` to adjust to unsupervised conditions. The grouping of these methods follows the same categorization as the supervised group above. Since a contemporaneous work, LLM4EA [22], is not yet open source, we only refer to its publicly available experimental results to ensure fairness in our comparisons.

To further enhance experimental comprehensiveness, we included three general methods involving LLM or multi-agent LLM ideas: (1) Zero-shot CoT [23]: a zero-shot Chain of Thought prompting method that enhances the effectiveness of answers through a step-by-step reasoning. (2) Self-Consistency [24]: A zero-shot CoT baseline that generates multiple reasoning paths and answers, selecting the most frequent answer as the final output. (3) Collaboration-Hard [25]: A collaboration method involving multiple LLM agents, where consensus determines the final answer.

Implementation details. All experiments are conducted on a server equipped with four NVIDIA GeForce RTX 4090 graphics cards, each featuring 24GB of GDDR6X memory. The system is powered by a 64-core processor and 480GB of system RAM. For storage, the server employs a 30GB system disk and a 50GB solid-state drive (SSD) for data storage. All experiments are implemented using the PyTorch framework. All baseline methods with * follow the same preprocessing procedure to obtain pseudo-labels for use as input. The LLMs in Table II all use the same settings, GPT-4. For subsequent experiments, unless otherwise stated, we employ GPT-3.5[3] as

[1] The codes are available at https://github.com/eduzrh/AdaCoAgentEA

[2] https://github.com/IDEA-FinAI/Simple-HHEA

[3] gpt-3.5-turbo-1106 from OpenAI API, https://openai.com/api/

TABLE I

Detailed statistics for the EA and HHEA datasets [1], [2]. "*Structure. Sim.*": The average neighbor structure similarity of entities, as defined in [1]. "*Temporal.*": Indicates whether the dataset includes temporal knowledge information. "Δ F.%", "Δ D.%": Indicates the relative difference in facts/density values between two KGs, using the KG with the smaller facts/lower density as the base.

Dataset		#Entities	#Relations	#Facts	Δ F.%	Density	Δ D.%	#Anchors	Temporal	Overlapping	Struc. Sim.
DBP15K(EN-FR) (Cross-lingual KGs)	EN	15,000	193	96,318	20.2%	6.421	20.2%	15,000	No	100%	63.4%
	FR	15,000	166	80,112		5.341			No	100%	
DBP-WIKI (HKGs)	DBP	100,000	413	293,990	16.8%	2.940	16.8%	100,000	No	100%	74.8%
	WIKI	100,000	261	251,708		2.517			No	100%	
ICEWS-WIKI (HHKGs)	ICEWS	11,047	272	3,527,881	1679.4%	319.352	2460.6%	5,058	Yes	45.79%	15.4%
	WIKI	15,896	226	198,257		12.472			Yes	31.82%	
ICEWS-YAGO (HHKGs)	ICEWS	26,863	272	4,192,555	3814.0%	156.072	3212.2%	18,824	Yes	70.07%	14.0%
	YAGO	22,734	41	107,118		4.712			Yes	82.80%	

the default LLM due to its cost efficiency.

In the Structural-Semantic Decoupling Gating, a degree difference threshold ($K_d = 50$) and structural similarity threshold ($\theta_{\text{str}} = 0.5$) are employed. S1 performs entity embedding by partitioning HHKG2 entity names into non-overlapping chunks (by rows, one row for each entity name), retrieving top-5 candidates for each batch of 10 queried HHKG1 entities. The S2 implements PageRank-based filtering with a 10% selection threshold. The S3 filtering entities through parameters $d_1 = 10$ and $d_2 = 3$. The S4 (using Simeple-HHEA) training with a learning rate of 0.01, weight decay of 0.001, gamma of 1.0, and 500 epochs. The $G3 = 0$, if the Hits@1 improvement in Area 2 falls below a 0.01 threshold compared to the prior phase, collectively balancing computational efficiency and model performance. The similarity threshold $\theta_{\text{u}} = 50\%$ in the output U_p of S4.

Evaluation metrics. Consistent with benchmark works [1], [3], we employ two widely established metrics to evaluate the performance of entity alignment: (1) Hits@k: This metric quantifies the percentage of correct alignments found within the top k ranked candidates (where $k = 1, 10$). Specifically, Hits@1 measures the proportion of exact matches, while Hits@10 captures the accuracy within the top 10 predictions. (2) Mean Reciprocal Rank (MRR): This metric reflects the average inverse ranking of correct results. For both metrics, higher values indicate superior entity alignment performance. Since some models only return final alignment results, Hits@1 is replaced with precision for performance evaluation.

References

[1] X. Jiang, C. Xu, Y. Shen, Y. Wang, F. Su, Z. Shi, F. Sun, Z. Li, J. Guo, and H. Shen, "Toward practical entity alignment method design: Insights from new highly heterogeneous knowledge graph datasets," in *Proceedings of the ACM on Web Conference 2024, WWW 2024, Singapore, May 13-17, 2024*, T. Chua, C. Ngo, R. Kumar, H. W. Lauw, and R. K. Lee, Eds. ACM, 2024, pp. 2325–2336. [Online]. Available: https://doi.org/10.1145/3589334.3645720

[2] X. Jiang, Y. Shen, Z. Shi, C. Xu, W. Li, Z. Li, J. Guo, H. Shen, and Y. Wang, "Unlocking the power of large language models for entity alignment," in *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2024, Bangkok, Thailand, August 11-16, 2024*, L. Ku, A. Martins, and V. Srikumar, Eds. Association for Computational Linguistics, 2024, pp. 7566–7583. [Online]. Available: https://aclanthology.org/2024.acl-long.408

[3] Z. Sun, Q. Zhang, W. Hu, C. Wang, M. Chen, F. Akrami, and C. Li, "A benchmarking study of embedding-based entity alignment for knowledge graphs," *Proc. VLDB Endow.*, vol. 13, no. 11, pp. 2326–2340, 2020. [Online]. Available: http://www.vldb.org/pvldb/vol13/p2326-sun.pdf

[4] W. Zeng, J. Zhou, and X. Zhao, "Benchmarking challenges for temporal knowledge graph alignment," *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:273501043

[5] M. Chen, Y. Tian, M. Yang, and C. Zaniolo, "Multilingual knowledge graph embeddings for cross-lingual knowledge alignment," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, Aug. 2017.

[6] Z. Sun, W. Hu, Q. Zhang, and Y. Qu, "Bootstrapping entity alignment with knowledge graph embedding," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, vol. 18. International Joint Conferences on Artificial Intelligence Organization, Jul. 2018, pp. 4396–4402. [Online]. Available: https://doi.org/10.24963/ijcai.2018/611

[7] Z. Wang, Q. Lv, X. Lan, and Y. Zhang, "Cross-lingual knowledge graph alignment via graph convolutional networks," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2018, pp. 349–357. [Online]. Available: https://doi.org/10.18653/v1/d18-1032

[8] Z. Chen, Y. Wu, Y. Feng, and D. Zhao, "Integrating manifold knowledge for global entity linking with heterogeneous graphs," *Data Intelligence*, vol. 4, no. 1, pp. 20–40, 2022.

[9] X. Mao, W. Wang, Y. Wu, and M. Lan, "Boosting the speed of entity alignment 10 ×: Dual attention matching network with normalized hard sample mining," in *Proceedings of the Web Conference 2021*. ACM, Apr. 2021, pp. 821–832. [Online]. Available: https://doi.org/10.1145/3442381.3449897

[10] C. Xu, F. Su, and J. Lehmann, "Time-aware graph neural network for entity alignment between temporal knowledge graphs," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2021, pp. 8999–9010. [Online]. Available: https://doi.org/10.18653/v1/2021.emnlp-main.709

[11] C. Xu, F. Su, B. Xiong, and J. Lehmann, "Time-aware entity alignment using temporal relational attention," in *Proceedings of the ACM Web Conference 2022*. ACM, Apr. 2022, pp. 788–797. [Online]. Available: https://doi.org/10.1145/3485447.3511922

[12] L. Cai, X. Mao, M. Ma, H. Yuan, J. Zhu, and M. Lan, "A simple temporal information matching mechanism for entity alignment between temporal knowledge graphs," in *Proceedings of the 29th International Conference on Computational Linguistics, COLING 2022, Gyeongju, Republic of Korea, October 12-17, 2022*. International Committee on Computational Linguistics, 2022, pp. 2075–2086. [Online]. Available: https://aclanthology.org/2022.coling-1.181

[13] X. Liu, J. Wu, T. Li, L. Chen, and Y. Gao, "Unsupervised entity alignment for temporal knowledge graphs," in *Proceedings of the ACM Web Conference 2023, WWW 2023, Austin, TX, USA, 30 April 2023 - 4 May 2023*, Y. Ding, J. Tang, J. F. Sequeda, L. Aroyo, C. Castillo, and G. Houben, Eds. ACM, 2023, pp. 2528–2538. [Online]. Available: https://doi.org/10.1145/3543507.3583381

[14] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis,

Minnesota: Association for Computational Linguistics, 2019, pp. 4171–4186. [Online]. Available: https://aclanthology.org/N19-1423

[15] C. Wang, Z. Huang, Y. Wan, J. Wei, J. Zhao, and P. Wang, "FuAlign: Cross-lingual entity alignment via multi-view representation learning of fused knowledge graphs," *Inform. Fusion*, vol. 89, pp. 41–52, Jan. 2023. [Online]. Available: https://doi.org/10.1016/j.inffus.2022.08.002

[16] X. Tang, J. Zhang, B. Chen, Y. Yang, H. Chen, and C. Li, "BERT}-{INT: A {BERT}-based interaction model for knowledge graph alignment," *interactions*, vol. 100, p. e1, 2020.

[17] F. M. Suchanek, S. Abiteboul, and P. Senellart, "Paris: Probabilistic alignment of relations, instances, and schema," *Proceedings of the VLDB Endowment*, vol. 5, no. 3, 2011.

[18] M. Leone, S. Huber, A. Arora, A. García-Durán, and R. West, "A critical re-evaluation of neural methods for entity alignment," *Proc. VLDB Endow.*, vol. 15, no. 8, pp. 1712–1725, 2022. [Online]. Available: https://www.vldb.org/pvldb/vol15/p1712-arora.pdf

[19] Q. Zhang, Z. Sun, W. Hu, M. Chen, L. Guo, and Y. Qu, "Multi-view knowledge graph embedding for entity alignment," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, S. Kraus, Ed. ijcai.org, 2019, pp. 5429–5435. [Online]. Available: https://doi.org/10.24963/ijcai.2019/754

[20] X. Liu, H. Hong, X. Wang, Z. Chen, E. Kharlamov, Y. Dong, and J. Tang, "Selfkg: Self-supervised entity alignment in knowledge graphs," in *WWW '22: The ACM Web Conference 2022, Virtual Event, Lyon, France, April 25 - 29, 2022*, F. Laforest, R. Troncy, E. Simperl, D. Agarwal, A. Gionis, I. Herman, and L. Médini, Eds. ACM, 2022, pp. 860–870. [Online]. Available: https://doi.org/10.1145/3485447.3511945

[21] N. Huo, R. Cheng, B. Kao, W. Ning, N. A. H. Haldar, X. Li, J. Li, M. M. Najafi, T. Li, and G. Qu, "Zeroea: A zero-training entity alignment framework via pre-trained language model," *Proc. VLDB Endow.*, vol. 17, no. 7, pp. 1765–1774, 2024. [Online]. Available: https://www.vldb.org/pvldb/vol17/p1765-huo.pdf

[22] S. Chen, Q. Zhang, J. Dong, W. Hua, Q. Li, and X. Huang, "Entity alignment with noisy annotations from large language models," *ArXiv*, vol. abs/2405.16806, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:270062763

[23] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. H. Chi, Q. V. Le, and D. Zhou, "Chain-of-thought prompting elicits reasoning in large language models," in *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds., 2022. [Online]. Available: http://papers.nips.cc/paper_files/paper/2022/hash/9d5609613524ecf4f15af0f7b31abca4-Abstract-Conference.html

[24] X. Wang, J. Wei, D. Schuurmans, Q. V. Le, E. H. Chi, S. Narang, A. Chowdhery, and D. Zhou, "Self-consistency improves chain of thought reasoning in language models," in *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. [Online]. Available: https://openreview.net/pdf?id=1PL1NIMMrw

[25] K. Xiong, X. Ding, Y. Cao, T. Liu, and B. Qin, "Examining inter-consistency of large language models collaboration: An in-depth analysis via debate," in *Findings of the Association for Computational Linguistics: EMNLP 2023, Singapore, December 6-10, 2023*, H. Bouamor, J. Pino, and K. Bali, Eds. Association for Computational Linguistics, 2023, pp. 7572–7590. [Online]. Available: https://doi.org/10.18653/v1/2023.findings-emnlp.508