Towards Temporal Knowledge Graph Alignment in the Wild
Technical Report

***Abstract*—This technical report contains the dataset details and the full experimental setup of the paper "Towards Temporal Knowledge Graph Alignment in the Wild".**

I. Supplementary Information on Methodology

Algorithm 1: Overall Framework of HyDRA

Input : Source TKG G^s, Target TKG G^t
Output: Final alignment ϕ_f

1 Initialize final alignment set $\phi_f \leftarrow \emptyset$
2 Encode G^s and G^t into multi-granular representations
3 **while** *not converged and iterations* $< I_n$ **do**
4 Compute embeddings: $\{\boldsymbol{h}_n^{dw}\}$, $\{\boldsymbol{h}_n^{(g)}\}_{g\in\mathcal{G}}$, $\{\boldsymbol{h}_n^{\text{name}}\}$
5 Adaptive integration: generate $\boldsymbol{P}^{ada}$ and select top-k_1 pseudo pairs
6 Dual-channel projection: $\tilde{\mathbb{P}}^{e^{s,t}}_{time}, \tilde{\mathbb{P}}^{e^{s,t}}_{rel} \leftarrow$ Eq. (5)
7 Construct projection hypergraph $\mathcal{G}_p$ and memory bank $\mathcal{M}_p$
8 Retrieve top-k_2 projections: generate $\mathcal{G}_m = \left(\bigcup_{\ell=1}^{3}\mathcal{V}_\ell, \bigcup_{\ell=1}^{3}\mathcal{E}_\ell\right)$
9 **for** *each scale* $\ell \in \{1,2,3\}$ **do**
10 Intra-scale interaction: $\mathcal{G}_\ell^* \leftarrow \mathcal{I}_\ell(\mathcal{G}_\ell)$
11 Fusion reasoning: $\phi_\ell \leftarrow$ LLM-Select$(\mathcal{G}_\ell^*)$
12 Conflict detection: $\mathcal{D}, \{C(e_i^s)\} \leftarrow$ Eq. (14)
13 Resolve conflicts: $\phi_C \leftarrow$ LLM-Select$(\{C(e_i^s)\})$
14 Update $\phi_f \leftarrow \bigcap_{\ell=1}^{3}\phi_\ell \cup \phi_C$
15 **return** ϕ_f

A. The Workflow of HyDRA

The overall workflow of HyDRA is presented in Algorithm 1. It takes as input a source TKG G^s and a target TKG G^t, and outputs the final alignment ϕ_f. The algorithm initializes the alignment set as empty and encodes both TKGs into multi-granular representations (lines 1–2).

In the *encoding and integration stage* (lines 4–5), structural embeddings $\{\boldsymbol{h}_n^{dw}\}$, multi-granular temporal embeddings $\{\boldsymbol{h}_n^{(g)}\}_{g\in\mathcal{G}}$ where $\mathcal{G} = \{\text{year}, \text{month}, \text{date}\}$, and name embeddings $\{\boldsymbol{h}_n^{\text{name}}\}$ are computed (line 4). Adaptive integration generates the similarity matrix $\boldsymbol{P}^{ada}$ and selects top-k_1 pseudo-aligned pairs (line 5).

In the *projection and retrieval stage* (lines 6–8), dual-channel projections $\tilde{\mathbb{P}}^{e^{s,t}}_{time}$ and $\tilde{\mathbb{P}}^{e^{s,t}}_{rel}$ are computed via Eq. (5) (line 6). The projection hypergraph $\mathcal{G}_p$ and memory bank $\mathcal{M}_p$ are constructed (line 7). Top-k_2 projections are retrieved to generate the multi-scale hypergraph $\mathcal{G}_m = \left(\bigcup_{\ell=1}^{3}\mathcal{V}_\ell, \bigcup_{\ell=1}^{3}\mathcal{E}_\ell\right)$ (line 8).

In the *fusion stage* (lines 9–14), for each scale $\ell \in \{1,2,3\}$, intra-scale interaction refines $\mathcal{G}_\ell$ into $\mathcal{G}_\ell^*$ via LLM-guided supplementation and trimming (line 10), and fusion reasoning generates scale-specific alignments ϕ_ℓ (line 11). Conflicts $\mathcal{D}$ and conflict sets $\{C(e_i^s)\}$ are detected via Eq. (14) (line 12), and resolved through LLM selection to obtain ϕ_C (line 13). The final alignment is updated as $\phi_f = \bigcap_{\ell=1}^{3}\phi_\ell \cup \phi_C$ (line 14). After iterative refinement, the algorithm returns ϕ_f (line 15).

II. Dataset

In this section, we introduce several key metrics used to quantify the characteristics and differences between temporal knowledge graph alignment (TKGA) datasets for assessing the difficulty and realism.

Overlapping Rate (Overlapping.%). To better reflect the real-world scenario where cross-KG alignment is rarely strictly 1-to-1, we use the overlapping rate to measure the degree of shared temporal entities between two TKGs [1]. The overlapping rate is calculated as follows:

$$\textbf{Overlapping.\%}(G^s) = \frac{|\phi_{seed}|}{|E^s|} \times 100\%, \tag{1}$$

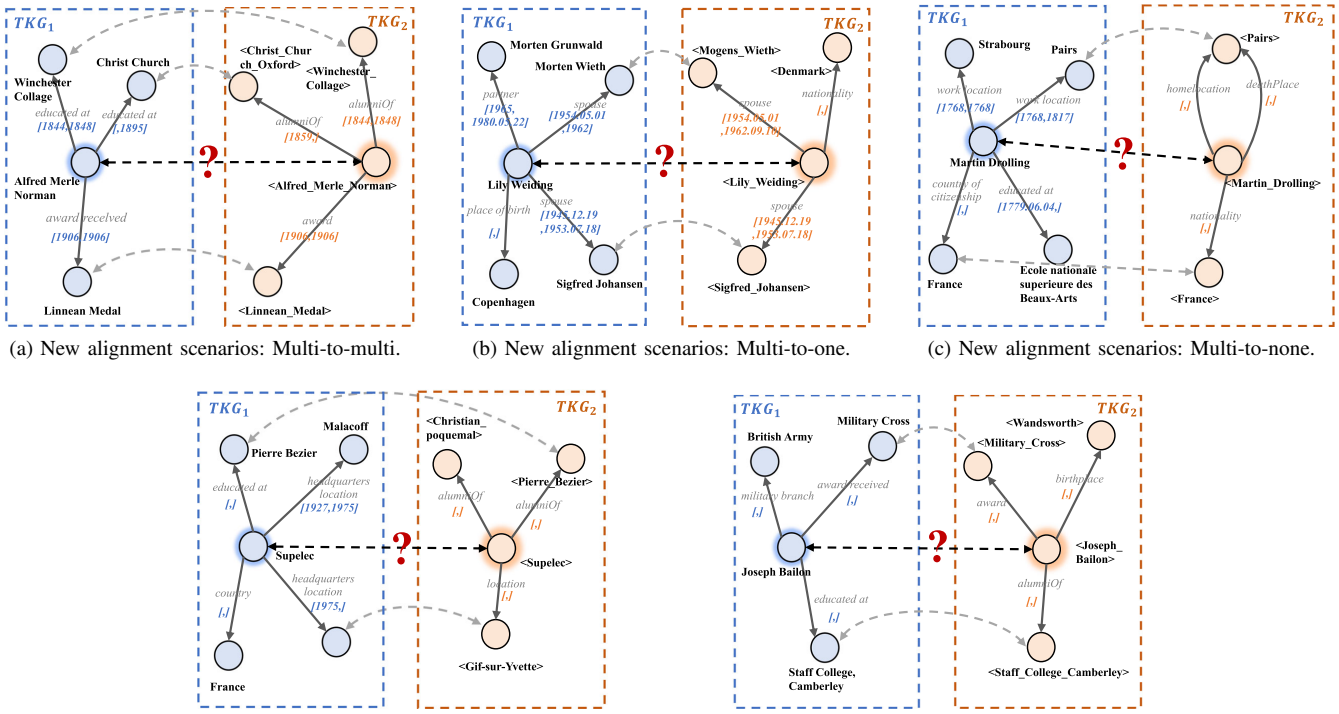
$$\textbf{Overlapping.\%}(G^t) = \frac{|\phi_{seed}|}{|E^t|} \times 100\%, \tag{2}$$

where ϕ_{seed} denotes the set of aligned entity pairs, and E^s and E^t are the set of all temporal entities in G^s and G^t.

Temporal Interval Consistency (Inter. Consis.%). To effectively assess the differences in temporal intervals between aligned entities in TKGA, we introduce a new metric called *temporal interval consistency*, which is the proportion of aligned entities with consistent temporal intervals among all aligned entities, defined as follows:

$$\textbf{Inter.Consis.\%}(G^s, G^t) = \\ = \frac{|(e_i^s, e_j^t) \in \phi_{seed} \mid \mathcal{T}(e_i^s) \cap \mathcal{T}(e_j^t) \neq \emptyset|}{|\phi_{seed}|} \times 100\%, \tag{3}$$

where ϕ_{seed} is the set of aligned entity pairs, and $\mathcal{T}(e)$ denotes the set of time intervals associated with entity e.



(a) New alignment scenarios: Multi-to-multi. (b) New alignment scenarios: Multi-to-one. (c) New alignment scenarios: Multi-to-none.

(d) New alignment scenarios: One-to-none. (e) New alignment scenarios: None-to-none.

Fig. 1: New alignment scenarios in BETA.

Multi-source Valid Temporal Fact Ratio (MTF.%). To assess the extent of missing valid temporal facts in multi-source TKGs, we define the multi-source valid temporal fact ratio as follows:

$$\textbf{MTF.\%}(G^s, G^t) = \frac{|\mathcal{V}G^s + \mathcal{V}G^t|}{|G^s + G^t|} \times 100\%, \quad (4)$$

where $|\mathcal{V}G^s + \mathcal{V}G^t|$ represents the total number of valid temporal facts in the source and target TKGs, and $|G^s + G^t|$ denotes the total number of all facts in both TKGs.

Differences in Valid Temporal Facts and Valid Temporal Density (ΔT.F.% and ΔT.D.%). To better capture the imbalance of valid temporal facts and temporal structures between two TKGs in real-world TKGA scenarios, we define the relative differences in valid temporal facts and valid temporal fact density, respectively, as follows:

$$\Delta\textbf{T.F.\%}(G^s, G^t) = \frac{|\mathcal{V}G^s - \mathcal{V}G^t|}{\min(|\mathcal{V}G^s|, |\mathcal{V}G^t|)} \times 100\%, \quad (5)$$

$$\Delta\textbf{T.D.\%}(G^s, G^t) = \frac{|\rho^s - \rho^t|}{\min(\rho^s, \rho^t)} \times 100\%, \quad (6)$$

where $|\mathcal{V}G^s|$ and $|\mathcal{V}G^t|$ represent the number of valid temporal facts in source TKG and targe TKG, and $\rho = \frac{|\mathcal{V}G|}{|G|}$ denotes the density of valid temporal facts in a TKG.

A. New Alignment Scenarios

For existing datasets, the TKGA algorithms merely need to align the football players and teams between two KGs, as the majority of relations are *member_of_sports_team* and *plays_for*. In contrast, in BETA and WildBETA, due to the introduction of more realistic quadruples, in addition to the current alignment scenario where the entities to be aligned are only associated with one specific temporal relation, there are new temporal alignment scenarios emerging:

Multi-to-multi/multi-to-one Alignment. In this setting, at least one of the entities to be aligned is associated with multiple temporal relations. In both cases, entities are involved in multiple temporal quadruples, and it is important to mine useful signals from the complex structure while reducing the noise introduced by heterogeneous relational and temporal information.

Multi-to-none/one-to-none Alignment. In this setting, at least one of the entities to be aligned is not associated with any temporal relations. In such cases, the key is to properly address discrepancies in temporal information and sufficiently exploit features modeled from other aspects.

None-to-none alignment. In this setting, neither entity is associated with any temporal relation. As a result, temporal entity alignment is reduced to the general entity alignment task.

We present specific cases of multi-to-multi, multi-to-one, multi-to-none, one-to-none, and none-to-none alignment regarding the temporal relations in Fig. 1.

TABLE I: Dataset statistics [1]–[5]. "*#Ent*", "*#Rel.*", "*#Facts*", "*#T.Facts*": The number of entities, relations, quadruples and quadruples with valid time interval in KG1 (KG2), respectively. "*Temp.*", "*Multi-Granular*" : Indicates whether the dataset includes temporal knowledge information and the dataset includes multi-granular temporal knowledge information, respectively. "*#Overlapping*": Represents the proportion of overlapping temporal entities in KG1 and KG2. "*Inter. Consis.*" : Represents the proportion of aligned entities with consistent temporal intervals among all aligned entities. "Δ *T.F.%*", "Δ *T.D.%*" : Relative difference in valid temporal facts/density values between two KGs, using the KG with the smaller valid temporal facts/lower valid temporal density as the base. "*Multi-Source*", "*MTF.%*" : Refers to whether both TKGs in the dataset are temporal incompleteness, and the average proportion of valid temporal facts in the two TKGs, respectively.

Dataset		#Ent.	#Rel.	Temp.	Multi-Granular	#Seed	#Overlapping	Inter. Consis. ⇓	#Facts	#T.Facts	Δ T.F.% ⇑	#T.Density	Δ T.D.% ⇑	Multi-Source	MTF.% ⇓
AGROLD	AGROLD	96,117	7	✗	✗	11,555	✗	✗	49,924	✗	✗	✗	✗	✗	✗
	AGROLD	51,488	4	✗	✗		✗		364,606	✗		✗			
DOREMUS	-	2,057	19	✗	✗	238	✗	✗	6,832	✗	✗	✗	✗	✗	✗
	-	1,889	20	✗	✗		✗		5,543	✗		✗			
DBP15K(EN-FR)	EN	15,000	193	✗	✗	15,000	✗	✗	96,318	✗	✗	✗	✗	✗	✗
	FR	15,000	166	✗	✗		✗		80,112	✗		✗			
DBP-WIKI	DBP	100,000	413	✗	✗	100,000	✗	✗	293,990	✗	✗	✗	✗	✗	✗
	WIKI	100,000	261	✗	✗		✗		251,708	✗		✗			
ICEWS-WIKI	ICEWS	11,047	272	✓	✗	5,058	45.79%	55.63%	3,527,881	3,527,881	6,817.1%	319.352	9,853.4%	✗	96.05%
	WIKI	15,896	226	✓	✗		31.82%		198,257	51,002		3.208			
ICEWS-YAGO	ICEWS	26,863	272	✓	✗	18,824	70.07%	7.39%	4,192,555	4,192,555	13,764.3%	156.072	11,633.3%	✗	98.21%
	YAGO	22,734	41	✓	✗		82.80%		107,118	30,240		1.330			
DICEWS	ICEWS	9,517	247	✓	✗	8,566	90.01%	95.09%	307,552	307,552	0%	32.316	0.2%	✗	100%
	ICEWS	9,537	246	✓	✗		89.82%		307,553	307,553		32.248			
YAGO-WIKI50K	YAGO	49,629	11	✓	✗	49,172	99.08%	93.63%	221,050	221,050	43.8%	4.454	45.0%	✗	100%
	WIKI	49,222	30	✓	✗		99.90%		317,814	317,814		6.457			
BETA	WIKI	42,666	257	✓	✓	40,364	94.60%	55.12%	199,879	104,774	49.9%	2.456	48.6%	✓	48.22%
	YAGO	42,297	45	✓	✓		95.43%		162,320	69,896		1.653			
WildBETA	WIKI	27,519	301	✓	✓	17,124	62.23%	**5.27%**	527,977	142,145	**14,001.7%**	5.165	**13,722.5%**	✓	**25.37%**
	YAGO	26,975	40	✓	✓		63.48%		36,283	1,008		0.037			

III. Experiments

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

A. Experimental Setting

Datasets. In our experiments, we conducted comprehensive evaluations on **10 datasets**, including BETA, WildBETA, and 8 current datasets (as detailed in Tab. I).

Among these, DICEWS and YAGO-WIKI50K are the most frequently used datasets for Temporal Knowledge Graph Alignment (TKGA), derived from ICEWS05-15, YAGO, and Wikidata. Specifically, ICEWS05-15 is constructed from the ICEWS dataset [6], which comprises political events annotated with specific dates, using a daily temporal resolution and covering the period from 2005 to 2015. Xu et al. [3] randomly partitioned the quadruples in ICEWS05-15 into two equally sized subsets, yielding the datasets DICEWS-200 (D200). The YAGO-WIKI50K datasets are similarly constructed by Xu et al. [3], who first selected the top 50,000 most frequent entities from a Wikidata subset (as extracted in [7]) and linked them to corresponding entities in YAGO. Temporal facts were then added to form two temporally enriched knowledge graphs. The resulting dataset YAGO-WIKI50K-1K contains 1000 seed entity pairs.

In contrast, ICEWS-WIKI and ICEWS-YAGO [1] represent new heterogeneous and temporal datasets, posing more realistic and challenging alignment scenarios. These datasets are characterized by significant discrepancies not only in the number of entities, relations, and triples. Furthermore, the number of seeds is not directly proportional to the total entity count, adding complexity to the temporal alignment task.

In addition, two standard non-temporal KGA datasets, DBP15K(EN-FR) and DBP-WIKI [4], are also included. DBP15K(EN-FR) focuses on cross-lingual alignment, while DBP-WIKI offers a large-scale benchmark for aligning heterogeneous KGs. Both datasets exhibit similar structural properties and high overlap (100%) in aligned entities, relations, and facts.

We also consider DOREMUS and AGROLD. DOREMUS [8] is a dataset specifically designed for the classical music domain. AGROLD [9], on the other hand, is a knowledge graph used for plant science. These datasets are included to test our model on domain-specific data outside of the general-purpose knowledge graphs mentioned above.

Baselines. Currently, no specific solutions exist for TKGA-Wild. To establish a comprehensive baseline, we introduced 28 SOTA and classic baseline methods for extensive comparison:

- MTransE [10], which introduces translation vectors to align entity embeddings across languages; and
- AlignE [11], which employs neural relation extraction to identify key relationships; and
- BootEA [11], which is one of the most competitive translation-based EA methods; and
- GCN-Align [12], which trains GCNs to embed entities of each language into a unified vector space; and
- MRAEA [13], which applies attention over local neighborhoods and relation-level meta-information; and
- RREA [14], which implements relational reflection transformations to generate relation-aware embeddings; and
- RDGCN [15], which leverages GCNs for modeling struc-

[1]The source codes and datasets of the previous work are available at https://github.com/DexterZeng/BETA. The source codes and datasets for this extended work will be released upon acceptance

tural information within knowledge graphs.

- Dual-AMN [16], which iointly captures intra-graph and cross-graph dependencies; and
- TEA-GNN [3], which treats timestamps as link attributes, using a time-aware attention mechanism to enrich entity and relation representations; and
- TREA [17], which enhances training using neighborhood aggregation and margin-based multi-class loss; and
- STEA [18], which utilizes a temporal dictionary to guide temporal alignment; and
- Dual-Match [19], which employs a temporal encoder for unsupervised layer-wise propagation; and
- MGTEA [20], which proposes a simple yet effective multi-granular approach for temporal alignment; and
- LightTEA [21], which is a lightweight TKGA model, though its temporal component yields limited improvements on existing datasets; and
- BERT [22], utilized as a pretrained language model to initialize entity embeddings using name-based features; and
- FuAlign [23], which incorporates auxiliary information to address KG heterogeneity; and
- BERT-INT [24], which combines BERT-based augmentation with auxiliary cues for improved alignment; and
- PARIS [25], which is a probabilistic iterative method capable of aligning entities without prior alignments; and
- NMN [26], which utilizes a neighborhood matching network to capture static KG's similarities in structures; and
- SDEA [27], which is a semantic-driven embedding learning framework that addresses the semantic gap between static KGs through multiple semantic enhancement strategies; and
- Attr-Int [28], which is a simple yet effective framework designed to handle static KG by integrating attribute-level information with structural embeddings; and
- DAEA [5], which enhances static KGA by employing multi-source domain adaptation to bridge the gap between various KG distributions; and
- Simple-HHEA [1], which is a representation learning-based approach tailored for aligning heterogeneous and temporal KGs; and
- ChatEA [2], which applies large language models with fine-tuning to perform advanced KG alignment; and
- HTEA [29], which employs frequency-based temporal embeddings to enhance alignment performance; and
- Naive RAG [30], [31], a basic LLM-based RAG approach that first retrieves relevant information based on a user query and then generates answers using the retrieved content; and
- Self-Consistency [32], a chain-of-thought baseline that produces multiple reasoning paths and selects the most frequent answer as the final output. In our implementation, we further enhance it by using the top-1 most similar entity from the similarity matrix produced by Simple-HHEA as a preprocessing step for the knowledge graph; and
- Self-RAG [33], a self-reflective RAG method aimed at improving the generation quality of LLMs.

Implementation details. All experiments were conducted on a server equipped with four NVIDIA GeForce RTX 4090 graphics cards, each with 24 GB of GDDR6X memory. The system features a 64-core processor and 480 GB of RAM. For storage, the server utilizes a 30 GB system disk alongside a 50 GB solid-state drive (SSD) for data storage. All implementations were carried out using the PyTorch framework.

The large language models (LLMs) reported in Tab. III and Tab. IV were evaluated under identical settings, employing GPT-4[2], except for ChatEA, which directly follows the results reported in its original paper. For subsequent experiments, unless otherwise specified, GPT-3.5[3] was adopted as the default LLM owing to its cost-effectiveness.

The multi-granular information encoders and the integrated training were configured with a learning rate of 0.01, a weight decay of 0.001, gamma set to 1.0, and were trained for 500 epochs. The training set proportions follow the settings used in prior work and are set as follows: `WildBETA` (2%), `YAGO-WIKI50K-1K` (2%), `DICEWS-200` (2.3%), `BETA` (10%) [20], `ICEWS-WIKI` (30%) [1], `ICEWS-YAGO` (30%) [1], `DBP15K(EN-FR)` (30%), and `DBP-WIKI` (30%).

Evaluation metrics. Consistent with prior benchmark studies [1], [4], we adopt two widely recognized evaluation metrics to assess the effectiveness of entity alignment models: Hits@k and Mean Reciprocal Rank (MRR).

1) Hits@k evaluates the proportion of correctly aligned entity pairs that appear among the top-k ranked candidates. Formally, let N denote the total number of reference (ground truth) alignments, and for each reference entity e_i, let rank_i denote the rank position of its correct counterpart in the candidate list. The Hits@k is defined as:

$$\text{Hits@}k = \frac{1}{N}\sum_{i=1}^{N}\mathbb{I}(\text{rank}_i \leq k), \tag{7}$$

where $\mathbb{I}(\cdot)$ is the indicator function, which returns 1 if the condition is true and 0 otherwise. In practice, Hits@1 reflects the strict accuracy of top-1 predictions, while Hits@10 provides insight into broader top-k retrieval performance.

2) Mean Reciprocal Rank (MRR) measures the average of the reciprocal ranks of the correct entities in the prediction lists. It is computed as:

$$\text{MRR} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\text{rank}_i}. \tag{8}$$

MRR captures both the presence and position of correct alignments, thereby emphasizing early correct retrieval.

Both metrics are positively oriented, meaning higher values indicate better alignment quality. Notably, in cases where models yield only the final alignment predictions (i.e., without ranked candidate lists), the Hits@1 score is substituted with standard precision.

[2]gpt-4-0125-preview from the OpenAI API, https://openai.com/api/
[3]gpt-3.5-turbo-1106 from the OpenAI API, https://openai.com/api/

B. Discussion on Cost and Environmental Considerations of LLM Usage

Recent advances in LLMs have significantly expanded the modeling capacity of knowledge-driven systems. However, their growing computational cost and environmental impact have also raised increasing concerns. In this section, we provide a supplementary discussion on the cost–performance trade-offs and environmental considerations of incorporating LLMs into the proposed TKGA-Wild framework.

Cost and Efficiency Considerations. Although HyDRA leverages LLMs in the Intra-Scale Interaction and Multi-Scale Fusion Reasoning modules, it is explicitly designed to minimize unnecessary LLM invocations. Unlike conventional LLM-based RAG pipelines that repeatedly query the LLM for large candidate sets, our framework restricts LLM usage to a small number of high-confidence candidates identified through multi-scale hypergraph retrieval. This design substantially reduces both token consumption and inference latency.

As reported in Sec. 5.5 of the main paper, HyDRA achieves a reduction of up to 99% in token usage and over 85% in execution time compared to representative LLM-based RAG baselines, while simultaneously delivering superior alignment accuracy. These results indicate that the performance gains of HyDRA do not rely on excessive LLM computation, but rather on effective retrieval, structured reasoning, and multi-scale aggregation.

Environmental Impact and Sustainable Deployment. From an environmental perspective, reduced token consumption and shorter inference time directly translate into lower energy usage and carbon emissions during deployment. While accurately quantifying the carbon footprint of proprietary LLMs remains challenging due to limited disclosure of infrastructure details, token usage and runtime are widely recognized as practical and informative proxies for environmental impact.

Our empirical results suggest that HyDRA significantly mitigates the environmental burden typically associated with LLM-based systems. Moreover, experiments in Sec. 5.6 demonstrate that the proposed framework maintains strong performance even when powered by lighter LLMs (e.g., GPT-3.5), highlighting its reduced dependence on very large models and its potential for more sustainable real-world deployment.

Overall, this supplementary analysis underscores that HyDRA strikes a favorable balance between effectiveness and efficiency. By tightly integrating retrieval, hypergraph reasoning, and selective LLM utilization, the proposed framework not only advances state-of-the-art performance on TKGA-Wild, but also aligns with emerging best practices for responsible and sustainable use of large language models.

C. Generalization on More Static KGA Scenarios

To verify the adaptability of HyDRA, we conduct additional experiments on two domain-specific datasets: DOREMUS (classical music) and AGROLD (plant science). These datasets, typically used for static or domain-adaptive alignment (e.g., DAEA [5]), lack the explicit temporal dynamics of TKGA-Wild. As shown in Tab. II, HyDRA consistently outperforms all baselines. Notably, HyDRA achieves a 59.4% relative improvement in Hits@1 on AGROLD. This superior performance demonstrates that our multi-scale hypergraph retrieval-augmented paradigm and *scale-weave synergy* mechanism are not only robust to temporal disparities but also highly effective in capturing complex structural dependencies and overcoming schema asymmetry in diverse, specialized domains.

TABLE II: More results on DOREMUS and AGROLD.

Methods	DOREMUS			AGROLD		
	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR
RDGCN	0.012	0.109	-	0.001	0.003	-
NMN	0.000	0.004	-	0.001	0.001	-
SDEA	0.387	0.560	0.461	0.001	0.001	0.001
BERT-INT	0.479	0.593	0.515	0.215	0.250	0.229
Attr-Int	0.487	0.765	0.587	0.143	0.204	0.167
DAEA	0.778	0.886	0.815	0.271	0.349	0.300
HyDRA (Ours)	**0.834**	**0.903**	**0.858**	**0.432**	**0.537**	**0.455**

TABLE III: Sensitivity analysis of the retrieval size k_2 on TKGA datasets.

k_2	WildBETA	BETA	YAGO-WIKI50K-1K	DICEWS-200
3	0.801	0.856	0.954	0.975
5	0.843	0.892	0.968	0.984
8	0.866	0.916	0.975	0.988
10	**0.871**	**0.924**	**0.978**	**0.989**
12	0.869	0.923	0.977	0.988
15	0.865	0.919	0.975	0.987
20	0.861	0.915	0.973	0.986

D. Sensitivity Analysis of the Retrieval Size k_2

We conduct a sensitivity analysis to study the effect of the retrieval size k_2 in the relevant entity search module. Tab. III reports the Hits@1 performance under different k_2 values on four TKGA datasets.

When k_2 is small, the performance is noticeably limited, indicating that retrieving only a few projected entities provides insufficient evidence for accurate alignment. As k_2 increases from 3 to 5, Hits@1 improves substantially across all datasets, demonstrating the benefit of incorporating multiple relevant projections.

The performance becomes stable when k_2 lies in the range of 8 to 12, where only marginal fluctuations are observed. This stability is attributed to the fact that our subsequent modules perform inter-scale interaction, multi-scale fusion reasoning, and conflict detection, which collectively mitigate the sensitivity of the model to the specific choice of retrieval size. This plateau suggests that the most informative projected entities have already been retrieved and the framework effectively filters or integrates candidates.

Further increasing k_2 beyond this range does not yield additional gains and may slightly degrade performance, likely due to the introduction of less relevant or noisy projections. Based on this observation, we set $k_2 = 10$ as a representative and robust choice in all experiments.

E. Scalability and Feasibility for Large Scale Knowledge Graph Alignment Task

Large scale knowledge graph alignment, involving graphs with millions of entities, is widely recognized as an independent research task, distinct from conventional KGA settings such as benchmark scale datasets and temporally evolving graphs including TKGA Wild. Existing large scale KGA systems do not attempt to perform direct alignment over the full knowledge graph due to prohibitive computational and memory requirements. Instead, they follow a decomposition based paradigm, where the original graph is partitioned into a set of smaller subgraphs, and standard KGA algorithms are applied independently within each subgraph.

This paradigm has been widely adopted in the literature. Representative examples include LIME [34] and ELsEA [**?**], which introduce dedicated preprocessing modules such as blocking, graph partitioning, and structure aware subgraph extraction to reduce the effective problem size. Crucially, the downstream KGA method is treated as a generic inference component, and scalability is primarily achieved through preprocessing rather than by altering the alignment algorithm itself.

Although large scale knowledge graph alignment is beyond the primary scope of this work, the proposed HyDRA framework is equally applicable under the same decomposition based setting. Our proposed HyDRA framework naturally fits into this established pipeline. Specifically, decomposition methods such as LIME [34] can be employed as a preprocessing step to generate candidate subgraphs from large knowledge graphs. HyDRA can then be applied within each subgraph without any modification. Since hypergraph construction in HyDRA is retrieval bounded and linear in the number of source entities within each subgraph, and the iterative update process operates locally, the overall computational complexity scales with the subgraph size rather than the size of the original graph.

Therefore, under the same assumptions adopted by prior large scale KGA systems, namely the availability of a graph decomposition or blocking module, HyDRA is equally applicable to million scale alignment scenarios. This analysis indicates that HyDRA does not impose additional scalability constraints beyond those already present in the state of the art, while offering improved reasoning capability and favorable cost performance trade offs in LLM assisted knowledge graph alignment.

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.

[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.

[3] 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, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*. Association for Computational Linguistics, 2021, pp. 8999–9010.

[4] 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.

[5] L. Yang, S. Zhou, J. Cheng, F. Zhang, J. Wan, S. Wang, and M. Lee, "DAEA: enhancing entity alignment in real-world knowledge graphs through multi-source domain adaptation," in *Proceedings of the 31st International Conference on Computational Linguistics, COLING 2025, Abu Dhabi, UAE, January 19-24, 2025*, O. Rambow, L. Wanner, M. Apidianaki, H. Al-Khalifa, B. D. Eugenio, and S. Schockaert, Eds. Association for Computational Linguistics, 2025, pp. 5890–5901.

[6] J. Lautenschlager, S. Shellman, and M. Ward, "ICEWS Event Aggregations," 2015.

[7] T. Lacroix, G. Obozinski, and N. Usunier, "Tensor decompositions for temporal knowledge base completion," in *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020.

[8] P. Lisena, M. Achichi, P. Choffé, C. Cecconi, K. Todorov, B. Jacquemin, and R. Troncy, "Improving (re-)usability of musical datasets: An overview of the DOREMUS project," *CoRR*, vol. abs/2405.03382, 2024.

[9] A. Venkatesan, G. Tagny, N. El Hassouni, I. Chentli, V. Guignon, C. Jonquet, M. Ruiz, and P. Larmande, "Agronomic linked data (agrold): a knowledge-based system to enable integrative biology in agronomy," *bioRxiv*, 2018.

[10] M. Chen, Y. Tian, M. Yang, and C. Zaniolo, "Multilingual knowledge graph embeddings for cross-lingual knowledge alignment," in *IJCAI*, 2017, pp. 1511–1517.

[11] 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, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*. ijcai.org, 2018, pp. 4396–4402.

[12] 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, Brussels, Belgium, October 31 - November 4, 2018*. Association for Computational Linguistics, 2018, pp. 349–357.

[13] X. Mao, W. Wang, H. Xu, M. Lan, and Y. Wu, "MRAEA: an efficient and robust entity alignment approach for cross-lingual knowledge graph," in *WSDM '20: The Thirteenth ACM International Conference on Web Search and Data Mining, Houston, TX, USA, February 3-7, 2020*. ACM, 2020, pp. 420–428.

[14] X. Mao, W. Wang, H. Xu, Y. Wu, and M. Lan, "Relational reflection entity alignment," in *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020*. ACM, 2020, pp. 1095–1104.

[15] 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.

[16] 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 *WWW '21: The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19-23, 2021*. ACM / IW3C2, 2021, pp. 821–832.

[17] C. Xu, F. Su, B. Xiong, and J. Lehmann, "Time-aware entity alignment using temporal relational attention," in *WWW '22: The ACM Web Conference 2022, Virtual Event, Lyon, France, April 25 - 29, 2022*. ACM, 2022, pp. 788–797.

[18] 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.

[19] 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.

[20] 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.

[21] L. Cai, X. Mao, Y. Xiao, C. Wu, and M. Lan, "An effective and efficient time-aware entity alignment framework via two-aspect three-view label propagation," in *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI 2023, 19th-25th August 2023, Macao, SAR, China*. ijcai.org, 2023, pp. 5021–5029.

[22] 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.

[23] 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.

[24] X. Tang, J. Zhang, B. Chen, Y. Yang, H. Chen, and C. Li, "BERT-INT: A bert-based interaction model for knowledge graph alignment," in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, C. Bessiere, Ed. ijcai.org, 2020, pp. 3174–3180.

[25] 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.

[26] Y. Wu, X. Liu, Y. Feng, Z. Wang, and D. Zhao, "Neighborhood matching network for entity alignment," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, D. Jurafsky, J. Chai, N. Schluter, and J. R. Tetreault, Eds. Association for Computational Linguistics, 2020, pp. 6477–6487.

[27] Z. Zhong, M. Zhang, J. Fan, and C. Dou, "Semantics driven embedding learning for effective entity alignment," in *38th IEEE International Conference on Data Engineering, ICDE 2022, Kuala Lumpur, Malaysia, May 9-12, 2022*. IEEE, 2022, pp. 2127–2140.

[28] L. Yang, J. Cheng, C. Xu, X. Wang, J. Li, and F. Zhang, "Attr-int: A simple and effective entity alignment framework for heterogeneous knowledge graphs," in *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2024, Seoul, Republic of Korea, April 14-19, 2024*. IEEE, 2024, pp. 6315–6319.

[29] J. Li, W. Hua, F. Jin, and X. Li, "HTEA: heterogeneity-aware embedding learning for temporal entity alignment," in *Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining, WSDM 2025, Hannover, Germany, March 10-14, 2025*, W. Nejdl, S. Auer, M. Cha, M. Moens, and M. Najork, Eds. ACM, 2025, pp. 982–990.

[30] Q. Zhang, S. Chen, Y. Bei, Z. Yuan, H. Zhou, Z. Hong, J. Dong, H. Chen, Y. Chang, and X. Huang, "A survey of graph retrieval-augmented generation for customized large language models," *CoRR*, vol. abs/2501.13958, 2025.

[31] X. Ma, Y. Gong, P. He, H. Zhao, and N. Duan, "Query rewriting in retrieval-augmented large language models," in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, EMNLP 2023, Singapore, December 6-10, 2023*, H. Bouamor, J. Pino, and K. Bali, Eds. Association for Computational Linguistics, 2023, pp. 5303–5315.

[32] 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.

[33] A. Asai, Z. Wu, Y. Wang, A. Sil, and H. Hajishirzi, "Self-rag: Learning to retrieve, generate, and critique through self-reflection," in *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net, 2024.

[34] W. Zeng, X. Zhao, X. Li, J. Tang, and W. Wang, "On entity alignment at scale," *VLDB J.*, vol. 31, no. 5, pp. 1009–1033, 2022.