HIERARCHICAL CONSISTENCY CONSTRAINT FOR ROBUSTER UNSUPERVISED GRAPH ALIGNMENT

Anonymous authors
Paper under double-blind review

ABSTRACT

Unsupervised graph alignment aims to identify corresponding node pairs across two different graphs. Existing approaches typically utilize either structural or attribute information to obtain graph representation where intra-graph distances can be measured. However, the intertwined challenges of learning robust graph representations and effectively exploiting them for precise alignment remain key bottlenecks. To mitigate these issues, we propose a hierarchical consistency constraint (HCC) model that enforces hierarchical constraints on both the representation learning and alignment phases. Specifically, the graph distance consistency module works with a regularization term to learn representation robust to noise. For alignment, we propose the Karush–Kuhn–Tucker (KKT) consistency-enhanced Gromov-Wasserstein optimal transport (KGWOT) to ensure that the solution adheres to the KKT conditions during iterations. The KGWOT-based calculation module sparsifies the solution, leading to more accurate and robust alignment. Extensive experiments and analysis conducted on synthetic and real-world datasets demonstrate the robustness and accuracy of the proposed HCC model.

1 INTRODUCTION

Graph alignment aims to identify corresponding nodes across graphs and has been widely applied to user identity linkage Xu & Fung (2025), knowledge graph alignment Wang et al. (2024b), and protein–protein interaction prediction Gao et al. (2023). When ground-truth correspondences are available, it can be formulated as a supervised learning task Wang et al. (2023); Li et al. (2023); Zhang et al. (2021). However, reliable real-world labels are scarce due to privacy concerns and annotation cost Wang et al. (2024a), motivating the development of unsupervised graph alignment.

Existing unsupervised graph alignment methods follow two paradigms. The first learns graph representations without supervision and compares source and target nodes in a shared space Gao et al. (2021), whose assumption is restrictive since graphs often differ in feature dimensions or structural properties. The second formulates graph alignment as a Gromov-Wasserstein optimal transport (GWOT) problem Peyré et al. (2016); Mémoli (2014; 2011), which leverages the intra-graph pairwise distance for alignment and has shown effectiveness for heterogeneous graph alignment Yan et al. (2025); Tang et al. (2023a); Li et al. (2022); Xu et al. (2019); Titouan et al. (2019) .

Nevertheless, conventional GWOT based methods rely on a single and hand-crafted cost matrix, e.g., the adjacency matrix Li et al. (2022); Xu et al. (2019), to measure graph discrepancies. The fixed representations are suboptimal and sensitive to perturbations. Although FusedGW Titouan et al. (2019) combines Wasserstein and Gromov-Wasserstein distances to mitigate this issue, it still assumes a fixed representation space and fails with heterogeneous or divergent representation. In this context, recent studies Yan et al. (2025); Tang et al. (2023a) reformulate the unsupervised graph alignment problem as the joint optimization of graph representations and GWOT for more robust alignment. However, these approaches still face two key challenges: (1) Suboptimal and noise-sensitive representations. (2) Unreliable alignments caused by dense transport couplings. The former stems from jointly optimizing graph embeddings under the guidance of instantaneous and often inaccurate solutions. The latter is due to the dense couplings solved via Sinkhorn Cuturi (2013), which conflicts with the inherently sparse nature of ground-truth alignment, where true correspondences are limited and can be overwhelmed by spurious matches.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

To address these challenges, we propose the Hierarchical Consistency Constraint (HCC) framework that imposes hierarchical constraints on both the representation learning and alignment phases. Specifically, the Graph Distance Consistency (GDC) module guides graph representations toward noise-resistant graph-level distances. For alignment, we propose the KKT Consistency-enhanced GWOT (KGWOT), which encourages the solution to satisfy KKT consistency through the Wasserstein Alternative Fixed-point Iteration (WAFI) and KKT Consistency Regularized Iteration Method (KRIM), thereby sparsifying the resulting couplings and improving matching reliability. Experiments on both synthetic and real-world datasets demonstrate the robustness and accuracy of HCC.

The contributions of this paper are summarized as follows: (1) We identify and analyze two key challenges in GWOT-based approaches and propose a HCC framework to achieve robust alignment, which imposes consistency constraints on both graph representation learning and the GWOT solution process. (2) We propose the GDC module with the graph distance consistency term for learning of robust representation and the KGWOT-based calculation module which includes the WAFI step and the KRIM step for more accurate and sparse solutions. (3) We extensively evaluate the robustness and accuracy of HCC on four synthetic datasets under various settings and two real-world datasets. Extensive analysis is conducted to reveal the contribution and property of each component.

2 Related Works

2.1 Unsupervised Graph Alignment

Unsupervised graph alignment aims to identify corresponding nodes across two graphs without known anchor pairs. Existing methods mainly follow two directions. One compares learned node representations across graphs Liu et al. (2023); Pei et al. (2022); Heimann et al. (2018); Wang et al. (2018), but its performance is often limited by graph heterogeneity. The other formulates graph alignment as a GWOT problem, which infers correspondences by comparing pairwise intra-graph distances Xu et al. (2019); Titouan et al. (2019). Representative works include GWD Xu et al. (2019), which constructs graph representations from topology, and FusedGW Titouan et al. (2019), which further incorporates both structural and attribute information. Subsequent studies improve alignment accuracy, optimization efficiency, and downstream performance Kawakita et al. (2024); Tang et al. (2023b); Carrasco et al. (2023); Li et al. (2022); Chen et al. (2020). Despite their success, most GWOT-based methods rely on simple or hand-crafted cost matrices, limiting their robustness and adaptability. Recent methods therefore reformulate GWOT as a joint optimization problem over both the transport plan and the learnable graph representation space. For example, SLOTAlign Tang et al. (2023a) adopts proximal alternating linearized minimization, while HLOT Yan et al. (2025) introduces hypergraph-based representations. In contrast, the HCC framework learns more robust representations while encouraging more accurate alignments through a novel optimization strategy.

2.2 Gromov-Wasserstein Optimal Transport

Optimal Transport (OT) measures the discrepancy between different distributions. Monge Monge (1781) modeled OT as the problem of optimal mapping between two distributions. Kantorovich Kantorovich (2006) further generalized it into a more applicable form, which we term as the Kantorovich problem in this paper. There are many variants of OT with the regularization term added on the learning objective Peyré et al. (2019); Cuturi (2013). But the unavailability of reliable inter-graph cost matrix hinders the application of OT on graph learning tasks, e.g., unsupervised graph alignment. The GWOT is designed to measure different representation spaces via comparing intra-graph pair-wise distance Mémoli (2011; 2014); Peyré et al. (2016), which serves as an effective tool for the analysis of graph data. Previous GWOT based graph alignment methods optimize iteratively via proximal term based strategy and the Sinkhorn Cuturi (2013) algorithm is used for each inner iteration. In this paper, we propose to mitigate the dense property of Sinkhorn based methods via KKT consistency constraint and thus achieve more sparse and accurate graph alignment.

3 Methodology

Notations. To start with, we introduce the key notations. An undirected attributed graph is represented as the quadruple $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{Z}, \ell)$, where $(\mathcal{V}, \mathcal{E}, \mathcal{Z})$ denotes the set of nodes, edges, and

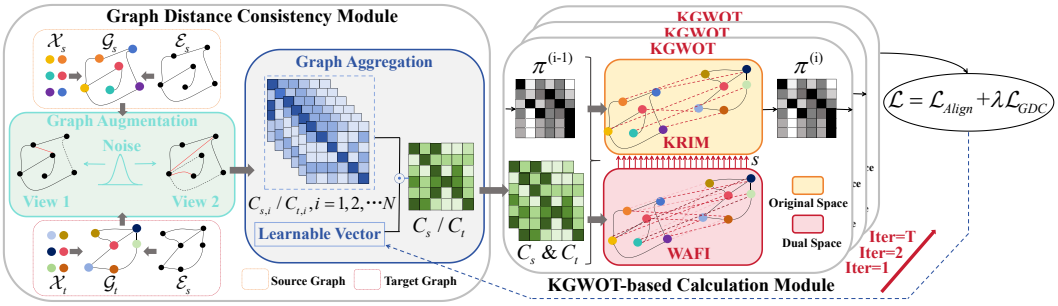


Figure 1: Model architecture diagram. The proposed HCC model consists of two modules, i.e., the GDC module and the KGWOT-based calculation module, where s is the solved dual variable via the WAFI step. The GDC module learns robust graph representation with $\mathcal{L}$ as the loss function.

attribution (i.e., the feature of the nodes) of $\mathcal{G}$. $\ell : \mathcal{V} \to \mathbf{\Omega}$ maps node $v_i \in \mathcal{V}$ to its representation $x_i \coloneqq \ell(v_i|\mathcal{E}, \mathcal{Z})$ under the conditions of $\mathcal{E}$ and $\mathcal{Z}$. The representation space defined on x_i is $(\mathbf{\Omega}, \mathbf{C})$, where $\mathbf{\Omega}$ is generally implicit and function $\mathbf{C} : \mathbf{\Omega} \times \mathbf{\Omega} \to \mathbb{R}_+$ measures intra-graph distance.

Problem Definition. In this paper, we focus on the unsupervised graph alignment task. For two undirected attributed graphs $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s, \mathcal{Z}_s, \ell_s)$ and $\mathcal{G}_t = (\mathcal{V}_t, \mathcal{E}_t, \mathcal{Z}_t, \ell_t)$, the optimal alignment $\mathcal{M}^*$ is to be solved via the following optimization problem:

$$\mathcal{M}^* = \arg\max_{\mathcal{M} \in \mathbb{M}} \mathcal{F}(\mathcal{G}_s, \mathcal{G}_t; \mathcal{M}), \tag{1}$$

where $\mathbb{M} = \{\mathcal{M} \in \{0,1\}^{|\mathcal{V}_s| \times |\mathcal{V}_t|} \mid \mathcal{M}\mathbf{1}_{|\mathcal{V}_t|} \leq \mathbf{1}_{|\mathcal{V}_s|}, \mathcal{M}^\top \mathbf{1}_{|\mathcal{V}_s|} \leq \mathbf{1}_{|\mathcal{V}_t|}\}$ is the set of all feasible alignments. The objective function $\mathcal{F}(\cdot)$ measures the affinity between the two graphs. In the unsupervised setting, $\mathcal{M}$ is learned without any ground-truth correspondence as supervision.

Model Architecture. To enhance the robustness and accuracy of unsupervised graph alignment, we propose the HCC model, which is a joint optimization framework composed of two main components, i.e., the GDC module and the KGWOT-based calculation module. The overall architecture of HCC is illustrated in Figure 1. Specifically, the GDC module aims at forcing a constraint on the learning process for more robust and noise-resistant representation. To overcome the inherent defect of inaccurate alignment caused by the dense solution of traditional GWOT, we propose the KGWOT-based calculation module, which includes two steps, i.e., WAFI and KRIM, to achieve accurate alignment. The HCC model leverages macro-level distance invariance with micro-level structural sparsification to provide a comprehensive solution to noise and alignment degradation.

3.1 Graph Distance Consistency Module

To improve unsupervised graph alignment under noisy settings, we propose the GDC module that encourages the learned representation to preserve stable inter-graph distances under perturbations. We first construct the intra-graph distance matrices by aggregating multiple sub-views:

$$\mathbf{C}_s = \sum_{k=1}^{K} \alpha_s^{(k)} \mathbf{C}_s^{(k)}, \quad \mathbf{C}_t = \sum_{k=1}^{K} \alpha_t^{(k)} \mathbf{C}_t^{(k)}, \tag{2}$$

where $\boldsymbol{\alpha}_s, \boldsymbol{\alpha}_t \in \mathbf{\Sigma}_K$ are learnable view weights, and $\mathbf{C}_s^{(k)}$, $\mathbf{C}_t^{(k)}$ denote the intra-graph measures from the k-th sub-view, such as edge-view, node-view, and subgraph-view. The construction details are provided in Appendix A.1. Given the learned representation space, the unsupervised graph alignment problem can be formulated as a generalized GWOT problem:

$$\min_{\boldsymbol{\pi} \geq 0, \boldsymbol{\alpha}_s, \boldsymbol{\alpha}_t} \langle \mathbf{L}(\mathbf{C}_s, \mathbf{C}_t, \boldsymbol{\pi}), \boldsymbol{\pi} \rangle, \quad \text{s.t. } \boldsymbol{\pi}\mathbf{1}_m = \boldsymbol{\mu}_s,\ \boldsymbol{\pi}^T \mathbf{1}_n = \boldsymbol{\mu}_t, \tag{3}$$

where $\boldsymbol{\mu}_s$ and $\boldsymbol{\mu}_t$ are the marginal distributions. $\mathbf{L}(\mathbf{C}_s, \mathbf{C}_t, \boldsymbol{\pi})$ denotes the Gromov-Wasserstein cost induced by the intra-graph distances. π_{ij} represents the matching score between $(v_{s,i}, v_{t,j})$, and the final alignment can be obtained by maximizing the matching score over all feasible matchings.

However, the quality of the representation space is highly sensitive to the view weights. Fixed or manually selected weights may lead to suboptimal performance, especially under noisy conditions.

Algorithm 1 Optimization Scheme for KGWOT

Require: Graph matrix $\mathbf{C}_s$, $\mathbf{C}_t$. Marginal probabilities $\boldsymbol{\mu}_s$, $\boldsymbol{\mu}_t$. Number of outer iterations T_{max}, number of inner iterations t_{max}, scaling factor γ, weighting factor ϵ.
Ensure: Optimal transport plan π^*.
1: Initialize $\boldsymbol{\pi}_0 = \boldsymbol{\mu}_s \boldsymbol{\mu}_t^T$.
2: **for** $1 \leq T \leq T_{max}$ **do**
3: Calculate the matrix $\bar{\mathbf{L}}^{(T)} = \mathbf{L}(\mathbf{C}_s, \mathbf{C}_t, \boldsymbol{\pi}^{(T)})$.
4: *# WAFI Step for KGWOT:*
5: **for** $1 \leq t \leq t_{max}$ **do**
6: Obtain dual variables $\boldsymbol{f}^{(t)}$ via Eq. (17).
7: **end for**
8: Obtain multiplier $\boldsymbol{s}^{(T)}$ via Eq. (11).
9: *# KRIM step for KGWOT:*
10: Obtain matching results $\boldsymbol{\pi}^{(T+1)}$ via Eq. (12).
11: **end for**
12: **return** Matching results of $\boldsymbol{\pi}^{(T_{max})}$.

Although existing methods setting these weights learnable and jointly update them with GWOT, the learned representations can still suffer from severe degradation when graph structures are perturbed.

To address this issue, we augment the graphs as $\widetilde{\mathcal{G}}_s = \mathcal{A}(\mathcal{G}_s)$ and $\widetilde{\mathcal{G}}_t = \mathcal{A}(\mathcal{G}_t)$, where $\mathcal{A}$ denotes graph augmentation, e.g., edge reconstruction or feature manipulation. We then evaluate whether the transport plan solved on the original graphs remains consistent on the augmented graphs with the GDC regularization term defined as follows:

$$\mathcal{L}_{GDC} = \left\langle \mathbf{L}(\widetilde{\mathbf{C}}_s, \widetilde{\mathbf{C}}_t, \boldsymbol{\pi}^*), \boldsymbol{\pi}^* \right\rangle, \tag{4}$$

where $\widetilde{\mathbf{C}}_s$ and $\widetilde{\mathbf{C}}_t$ are the intra-graph distance matrices of the augmented graphs, and $\boldsymbol{\pi}^*$ is the optimal transport plan obtained from the original graphs. This term encourages the learned representation to produce stable graph-level distances under perturbations. The overall objective is:

$$\mathcal{L} = \mathcal{L}_{Align} + \lambda \mathcal{L}_{GDC} = \mathcal{J}_{GW}(\mathcal{G}_s, \mathcal{G}_t) + \lambda \mathcal{J}_{GW}(\widetilde{\mathcal{G}}_s, \widetilde{\mathcal{G}}_t), \tag{5}$$

where λ balances the alignment loss and the GDC regularization. In this way, the GDC module guides the model to learn more robust representations by enforcing graph distance consistency between original and augmented graphs. Since the optimization of the transport plan and the learning of representations are coupled, inaccurate transport solutions may impair representation learning. Entropy-regularized OT methods often produce dense and biased plans, resulting in less reliable distance estimates. Therefore, we propose the KGWOT-based calculation module, incorporating KKT consistency constraints to alleviate dense-solution and obtain more accurate alignment.

3.2 KGWOT-based Calculation Module

Given the representation spaces learned via the GDC module, we should subsequently solve the corresponding optimal transport plan. Previous approaches tend to solve the approximated solution via introducing entropic regularization. However, the introduced entropic regularization forces $\pi_{ij} > 0$, and thus the KKT condition no longer holds, which often leads to dense and inaccurate solutions. To this end, the KGWOT-based calculation module is proposed to obtain more accurate alignment from the perspective of KKT consistency, which is composed of the WAFI step and the KRIM step. Specifically, we first perform the WAFI step for the extraction of the dual variable related to KKT consistency. Then the KRIM step makes use of the extracted dual variable for the solution of the optimal transport plan. Details of the WAFI and KRIM step are illustrated as follows.

3.2.1 Wasserstein Alternative Fixed-point Iteration.

With fixed representation spaces of $\mathcal{G}_s$ and $\mathcal{G}_t$, the generalized GWOT problem in Eq. (3) reduces to the standard GWOT problem. However the resulting formulation is still non-convex. Thus, we follow common practice Yan et al. (2025) and update $\boldsymbol{\pi}$ via a proximal step as follows:

$$\boldsymbol{\pi}^{(T+1)} = \arg\min_{\boldsymbol{\pi}} \left\langle \mathbf{L}(\mathbf{C}_s, \mathbf{C}_t, \boldsymbol{\pi}^{(T)}), \boldsymbol{\pi} \right\rangle, \tag{6}$$

Algorithm 2 Hierarchical Consistency Constraint Framework

Require: Graph $\mathcal{G}_s$ and $\mathcal{G}_t$; Graph augmentation operator $\mathcal{A}(\cdot)$; Weighting factor λ; Learning rate η; Max iterations I_{max}; Parameters for KGWOT: γ, ϵ, T_{max}.
Ensure: Optimal alignment matrix $\mathcal{M}^*$.

```
Initialize view weights α_s^(0), α_t^(0) and transport plan π^(0).
Compute initial intra-graph distances C_s and C_t via Eq. (2).
for i = 1 to I_max do
    # Phase 1: The GDC Module
    Obtain augmented graphs G̃_s = A(G_s), G̃_t = A(G_t).
    Calculate the GDC-regularized loss L using current π^(i−1) as defined in Eq. (5).
    Update weights α_s and α_t via gradient descent based on L.
    Update representation spaces C_s^(i) and C_t^(i) via Eq. (2).
    # Phase 2: KGWOT-based Calculation Module
    Fix the representation spaces C_s^(i) and C_t^(i).
    Obtain the refined transport plan π^(i) via Algorithm 1.
end for
# Final Alignment Mapping
Solve for the discrete optimal alignment M* based on the final plan π^(I_max) via Eq. (1).
return optimal alignment M*.
```

where T denotes the current outer iteration number and $\boldsymbol{\pi}^{(T)}$ is fixed as the transport plan obtained at the T-th iteration. For simplicity, we let $\bar{\mathbf{L}}^{(T)} = \mathbf{L}(\mathbf{C}_s, \mathbf{C}_t, \boldsymbol{\pi}^{(T)})$. The dual form of Eq. (6) is:

$$\max_{\boldsymbol{f},\boldsymbol{g},\boldsymbol{s}} \mathcal{J} = \langle \boldsymbol{f}, \boldsymbol{\mu}_s \rangle + \langle \boldsymbol{g}, \boldsymbol{\mu}_t \rangle, \quad s.t.\ f_i + g_i + s_{ij} = \bar{L}_{ij}^{(T)}, s_{ij} \geq 0, \tag{7}$$

where $\boldsymbol{f}$, $\boldsymbol{g}$, and $\boldsymbol{s}$ are dual variables. To further extract the value of these dual variables, we adopt the following lemma to simplify the problem accordingly.

Lemma 1. *Given dual formulation in Eq. (7), we replace $\boldsymbol{g}$ via $g_j = \inf_{k\in[M]} \left(\bar{L}_{kj}^{(T)} - f_k\right)$ Villani et al. (2009), which yields the following unconstrained formulation:*

$$\min_{\boldsymbol{f}} \mathcal{J}_{\mathrm{U}} = \sum_{j=1}^{N} \mu_{t,j} \sup_{k\in[M]} \left(f_k - \bar{L}_{kj}^{(T)}\right) - \sum_{i=1}^{M} f_i \mu_{s,i}. \tag{8}$$

Although $\mathcal{J}_{\mathrm{U}}$ is a convex objective without explicit constraints, the inclusion of the $\sup(\cdot)$ operator introduces non-smoothness, which poses a challenge for efficient numerical optimization. To facilitate a faster optimization process, we adopt a smooth approximation by replacing the $\sup(\cdot)$ operator with the $\mathrm{LogSumExp}(\cdot)$ function, following the framework established in Nesterov (2005). Consequently, we derive the approximate Fenchel-Lagrange dual form of OT with given ϵ as follows:

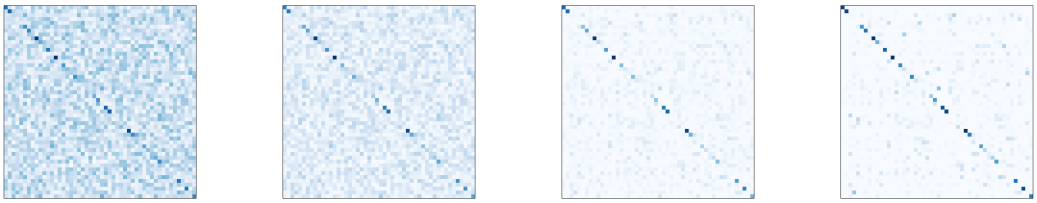
$$\min_{\boldsymbol{f}} \mathcal{J}_{FL} = \sum_{j=1}^{N} \left\{ \epsilon \log \left[\sum_{i=1}^{M} \exp\left(\frac{f_i - \bar{L}_{ij}^{(T)}}{\epsilon} \right) \right] \right\} \cdot \mu_{t,j} - \sum_{i=1}^{M} f_i \mu_{s,i}. \tag{9}$$

Notably, in this formulation, only the potential vector $\boldsymbol{f}$ needs to be optimized without any additional constraints. For efficient computing, we propose to optimize $\boldsymbol{f}$ via the following WAFI approach.

Proposition 1 (The iteration method of WAFI)**.** *For the problem of Eq. (9), we leverage the fixed-point method and obtain a new iteration formula as follows with $\mathcal{K}(f_j) = \sum_{k=1}^{M} \exp\left(\frac{f_k - \bar{L}_{kj}^{(T)}}{\epsilon}\right)$:*

$$f_i^{(t+1)} = \epsilon \left[\log(\mu_{s,i}) - \log \sum_{j=1}^{N} \left[\mu_{t,j} \exp\left(-\frac{\bar{L}_{ij}^{(T)}}{\epsilon} \right) \mathcal{K}^{-1}\left(f_j^{(t)}\right) \right] \right]. \tag{10}$$

Given that only the dual variable $\boldsymbol{f}$ need to be optimized, the WAFI approach reduces the time complexity by half compared to entropic regularization based methods. Please refer to Appendix B



(a) GWOT (b) KGWOT, $\gamma = 1$ (c) KGWOT, $\gamma = 5$ (d) KGWOT, $\gamma = 10$

Figure 2: Visualization of the optimal transport plans. The diagonal represents the true labels.

for detailed illustrations and the convergence proof. Then the dual variable s can be extracted as:

$$s_{ij} = \bar{L}_{ij}^{(T)} - f_i + \epsilon \log \left[\sum_{k=1}^{M} \exp \left(\frac{f_k - \bar{L}_{kj}^{(T)}}{\epsilon} \right) \right] \geq 0. \tag{11}$$

Through the proposed WAFI step, the dual variable $\boldsymbol{s}$ can be obtained efficiently. We subsequently leverage $\boldsymbol{s}$ for the accurate solution of the transport plan via the following KRIM step.

3.2.2 KKT Consistency Regularized Iteration Method.

Given the obtained $\boldsymbol{s}$, we should make use of it for the solution of optimal transport plan. We propose the KRIM step for the accurate and sparse solution of $\boldsymbol{\pi}$. Specifically, we incorporate $\boldsymbol{s}$ back into Eq. (6) as the regularization for KKT consistency constraint:

$$\min_{\boldsymbol{\pi} \geq 0} \langle \boldsymbol{\pi}, \mathbf{L} \odot \gamma \boldsymbol{s} \rangle - \epsilon \langle \boldsymbol{\pi}, \log(\boldsymbol{\pi}) - 1 \rangle, \tag{12}$$

where γ is a scaling factor. This formulation encourages entries π_{ij} associated with large s_{ij} values to approach zero, thereby promoting sparsity in the alignment prediction and maintaining consistency with the KKT condition. The dual Fenchel-Lagrange form of Eq. (12) is:

$$\max_{\boldsymbol{\psi}, \boldsymbol{\phi}} \mathcal{J}_{\mathrm{T}} = \boldsymbol{\mu}^{s\top} \boldsymbol{\psi} + \boldsymbol{\mu}^{t\top} \boldsymbol{\phi} - \epsilon \sum_{i,j} \exp \left(\frac{\psi_i + \phi_j - \mathscr{C}_{ij}}{\epsilon} \right). \tag{13}$$

where $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ denote the multipliers. $\mathscr{C}_{ij} = \gamma \bar{L}_{ij} s_{ij}$ denotes the reweighted cost matrix and the matching results $\boldsymbol{\pi}$ can be obtained:

$$\pi_{ij} = \exp \left(\frac{\psi_i + \phi_j - \mathscr{C}_{ij}}{\epsilon} \right) = \exp \left(\frac{\psi_i + \phi_j - \gamma \bar{L}_{ij} s_{ij}}{\epsilon} \right). \tag{14}$$

The dual potentials $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ appearing in Eq.(13) can be obtained via the Sinkhorn algorithm Cuturi (2013) with the following iteration:

$$\begin{cases} \exp \left(\dfrac{\phi_j^{(t+1)}}{\epsilon} \right) = \dfrac{\mu_{t,j}}{\sum\limits_{i=1}^{M} \exp \left(\frac{\psi_i^{(t)}}{\epsilon} \right) \exp \left(-\frac{\mathscr{C}_{ij}}{\epsilon} \right)} \\ \exp \left(\dfrac{\psi_i^{(t+1)}}{\epsilon} \right) = \dfrac{\mu_{s,i}}{\sum\limits_{j=1}^{N} \exp \left(\frac{\phi_j^{(t+1)}}{\epsilon} \right) \exp \left(-\frac{\mathscr{C}_{ij}}{\epsilon} \right)} \end{cases}, \tag{15}$$

where t denotes the current inner iteration number. The detailed scheme is shown in Algorithm 1. The KKT consistency constraint serves as a plug-and-play approach that ensures the accuracy while maintaining the time complexity. Figure 2 visualizes the optimal transport plans obtained via KGWOT compared to that obtained via vanilla GWOT solver, which employs Sinkhorn algorithm to update $\boldsymbol{\pi}$ iteratively. The vanilla GWOT solver yields a dense transport plan where the true matching being obscured by noisy and incorrect matches. In contrast, KGWOT assigns higher scores to true correspondences, resulting in sparser $\boldsymbol{\pi}$. As the scaling factor γ increases, $\boldsymbol{\pi}$ becomes sparser.

Summary. To mitigate the two challenges faced with graph alignment, i.e., (1) how to learn robust representations to counteract the effects of noise and (2) how to accurately solve for the optimal transmission plan in the learned representation space, the HCC model leverages the GDC module for the learning of more robust graph representation and adopts the KGWOT-based calculation module to suppress dense solutions and obtain the final alignment. Details of HCC are in Algorithm 2.

Table 1: The performance comparison on synthetic datasets, i.e., Cora, Citeseer, PPI, and Facebook. We report the Hit@1 accuracy under three different difficulty settings: Basic (30% perturbation noise), Medium (60% perturbation noise), and Hard (90% perturbation noise).

Dataset	Cora			Citeseer			PPI			Facebook		
Setting	Basic	Medium	Hard	Basic	Medium	Hard	Basic	Medium	Hard	Basic	Medium	Hard
REGAL	26.76	17.06	13.3	8.77	7.54	7.33	16.48	6.82	5.11	45.76	30.51	20.42
GCNAlign	49.92	20.34	10.17	39.03	15.23	6.05	40.35	16.6	10.81	0.02	0	0
GATAlign	48.52	19.25	8.14	37.88	14.85	5.92	21.32	4.35	5.22	3.69	0	0
WAlign	66.51	39.91	29.11	48.52	25.10	15.67	55.19	32.95	19.31	66.22	49.36	39.38
GWD	44.44	19.41	8.14	17.24	1.13	0.02	80.25	63.7	3.81	90.94	68.90	0
FusedGW	91.24	74.18	44.91	71.98	52.8	28.49	82.83	74.24	28.34	98.52	96.14	75.36
SLOTAlign	93.61	82.64	59.08	**84.55**	69.31	44.27	86.25	78.49	19.64	**98.54**	96.43	76.06
HLOT	93.24	82.98	61.03	84.67	70.75	46.32	86.21	78.54	21.83	98.49	96.30	75.48
HCC	**93.87**	**83.46**	**64.18**	84.45	**71.75**	**51.31**	**87.04**	**78.66**	**38.77**	98.51	**96.76**	**76.70**

Table 2: The performance of methods on real-world datasets Douban and ACM-DBLP. Since the case in Douban and ACM-DBLP is more complicated, we report the Hit@1, Hit@5, Hit@10 and Hit@30 accuracy for comprehensive evaluation. The best results are highlighted in bold.

Dataset	Douban				ACM-DBLP			
Accuracy	Hit@1	Hit@5	Hit@10	Hit@30	Hit@1	Hit@5	Hit@10	Hit@30
KNN	3.31	10.38	16.64	30.05	49.25	59.46	63.42	69.61
REGAL	30.32	54.83	–	–	34.09	46.58	51.35	56.34
GCNAlign	20.93	34.44	39.62	50.72	38.43	68.46	77.64	86.89
GATAlign	23.70	36.94	44.01	57.16	14.21	34.07	42.12	49.00
WAlign	35.69	57.87	69.05	83.09	50.61	72.87	80.84	89.47
GWD	3.04	7.96	9.21	11.90	56.24	77.14	82.20	84.92
FusedGW	29.61	62.79	66.46	68.07	30.80	38.39	39.26	39.60
SLOTAlign	51.43	73.43	77.73	82.02	66.04	84.06	87.95	90.32
HLOT	57.33	79.70	84.17	87.12	65.94	84.70	88.76	91.32
HCC	**59.66**	**80.05**	**86.31**	**90.16**	**68.30**	**85.23**	**88.93**	**91.37**

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

Datasets. To evaluate the performance of HCC in terms of both robustness and accuracy, we conduct experiments on four synthetic datasets, inlcuding Cora Sen et al. (2008), Citeseer Sen et al. (2008),PPI Zitnik & Leskovec (2017), and Facebook Leskovec & Mcauley (2012) under various difficulty settings. Specifically, on each dataset, we introduce varying degrees of structural perturbation noise to evaluate the model's robustness against structural disturbances. Since existing methods have already achieved strong robustness to feature perturbations, this paper focuses on analyzing the impact of structural perturbations. Additionally, we assess the real-world graph alignment capability using Douban Online-Offline Zhang & Tong (2016) and ACM-DBLP Zhang & Tong (2018).

Evaluation Metrics. The Hit@k metric is employed to evaluate graph alignment accuracy, measuring the percentage of source graph nodes whose true correspondences in the target graph appear among the top-k predicted candidates. All ground-truth node pairs are utilized for evaluation.

Baselines. We benchmark HCC against a diverse suite of seven unsupervised graph alignment methods, each representing a distinct technical approach, including REGAL Heimann et al. (2018), GCNAlign Wang et al. (2018), GATAlign Veličković et al. (2017), WAlign Gao et al. (2021), GWD Xu et al. (2019), FusedGW Titouan et al. (2019), SLOTAlign Tang et al. (2023a), HLOT Yan et al. (2025). WAlign Gao et al. (2021), GWD Xu et al. (2019), FusedGW Titouan et al. (2019) are baselines based on OT or GWOT, while SLOTAlign Tang et al. (2023a) and HLOT Yan et al. (2025) formulate unsupervised graph alignment as a joint optimization problem. Detailed description of the baseline can be found in Appendix A.3.

Implementation Details. The optimization of HCC follows a two-stage procedure. In the first stage, graph representations and the transport plan are jointly optimized for 100 epochs. The second stage further refines π with fixed representations. For experiments on the four synthetic datasets, the learning rate for unsupervised learning of α_s and α_t is set to $lr = 0.001$, and λ is set to 0.001. The

Table 3: Ablation experiments on Cora and PPI. We report the performance with different configurations and difficulty settings.

Dataset	GDC	KKT	Basic	Medium	Hard
Cora	×	×	93.61	82.64	59.08
	×	✓	93.72	83.20	63.55
	✓	×	93.80	82.46	59.23
	✓	✓	**93.87**	**83.46**	**64.18**
PPI	×	×	86.25	78.49	19.64
	×	✓	86.42	78.78	38.48
	✓	×	86.64	78.83	29.43
	✓	✓	**87.04**	**78.86**	**38.77**

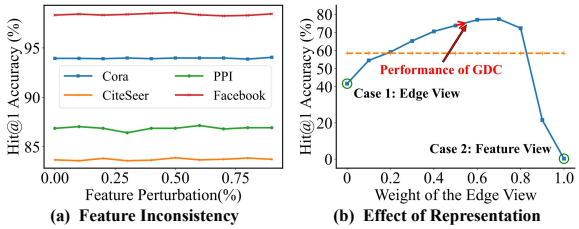


Figure 3: (a) Feature inconsistency experiments on four synthetic datasets. (b) The impact of graph representation on the alignment accuracy of the model.

Table 4: Empirical scaling comparison of runtime (s) between KGWOT and GWOT baseline.

N	**100**	**200**	**300**	**400**	**500**	**600**	**700**	**800**	**900**	**1000**
Entropy regularization	0.048	0.074	0.063	0.178	0.269	0.483	0.314	0.729	1.267	1.036
KGWOT	0.042	0.060	0.080	0.203	0.259	0.472	0.585	0.814	1.063	1.265

weight of the entropy term ϵ in Eq. (12) is set to $\epsilon = 0.005$, and the scaling factor is set to $\gamma = 1$, unless otherwise specified. The second-stage optimization runs for 1000 epochs. For real-world dataset experiments, we set $lr = 1$. We adopt a parameter-free GNN derived from the simplified GNN Wu et al. (2019) with both the parameterized linear layer and the activation function removed.

4.2 Main Results

Robustness Analysis. On four synthetic datasets, we report the Hit@1 accuracy under different perturbation settings in Table 1. For the basic, medium, and hard settings, we randomly redirect 30%, 60%, and 90% of edges in $\mathcal{G}_t$ to previously unconnected positions, respectively. To prevent models from relying solely on $\mathcal{Z}_s$ and $\mathcal{Z}_t$ for alignment, we retain only the first 100 feature columns of $\mathcal{G}_s$ and $\mathcal{G}_t$ on Cora, Citeseer, and Facebook. HCC consistently outperforms all baseline methods across all four datasets. As the perturbation intensity increases, HCC's advantage becomes more pronounced, demonstrating enhanced robustness attributable to its GDC and KGWOT modules. Specifically, we observe Hit@1 accuracy improvements of 5.1% on Cora, 7.04% on Citeseer, and 19.13% on PPI. The performance gain on Facebook, though positive, is relatively modest in comparison. The results verify the advantage of HCC in terms of accuracy and robustness.

Real-World Performance. To evaluate the real-world performance of different methods, we present Hit@1, Hit@5, Hit@10, and Hit@30 accuracy results on Douban and ACM-DBLP datasets in Table 2. HCC achieves state-of-the-art performance on both datasets across all four metrics. The consistent improvements observed not only in Hit@1 but also in Hit@5, Hit@10, and Hit@30 accuracy demonstrate the overall superiority of HCC. The experiments on both synthetic and real-world datasets demonstrate the robustness and the accuracy of HCC compared to the baselines. In the following subsection, we conduct ablation study and additional analysis about HCC.

4.3 Ablation Study and Further Analysis

Effectiveness of GDC and KGWOT. Table 3 presents the ablation results. All HCC variants outperform the baseline method, and the complete HCC framework achieves the best performance across all difficulty levels. The benefit of GDC is particularly evident in challenging alignment settings. For example, on PPI-Hard without KKT consistency, adding GDC improves alignment accuracy from 19.6 to 29.4. When KKT consistency is enabled, adding GDC still improves performance across all reported datasets and settings. These results suggest that GDC provides complementary benefits to KKT consistency, with a more pronounced contribution under noisy and challenging alignment conditions. Figure 3(b) further illustrates how the alignment accuracy of $\mathcal{M}$ varies with the view weights $[\boldsymbol{\alpha}_s^{(q)}]$ and $[\boldsymbol{\alpha}_t^{(q)}]$. The blue curve shows the performance of fixed-weight alignment under different weight configurations, while the orange line represents the KNN baseline. The red arrows connect the graph representations learned using only $\mathcal{L}_{Align}$ to those learned with additional $\mathcal{L}_{GDC}$

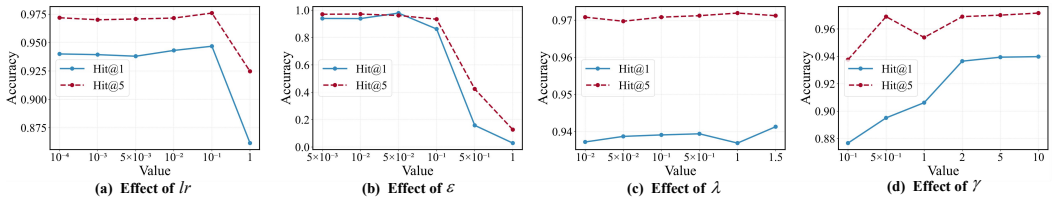


Figure 4: Hyperparameters sensitivity analysis. (a) The effect of the learning rate lr. (b) The effect of the weight of the entropy term ϵ. (c) The effect of the weight of graph distance consistency λ. (d) The effect scaling factor of γ. We report the Hit@1 accuracy (blue) and the Hit@5 accuracy (red).

guidance. Different views contribute unequally to the alignment task, making the choice of view weights important. Manually specified weights can lead to suboptimal performance, as illustrated by uniform weighting, or even poorer results, as in Cases 1 and 2. Although previous methods Yan et al. (2025); Tang et al. (2023a) employ projected proximal gradient updates to optimize $[\boldsymbol{\alpha}_s^{(q)}]$ and $[\boldsymbol{\alpha}_t^{(q)}]$ under the guidance of the GW distance, the performance still degrade in noisy settings.

Efficiency Analysis. HCC retains the same asymptotic computational order as the underlying GWOT framework. Specifically, WAFI requires only one dual-potential vector, avoiding the alternating computation of two dual potentials in each inner iteration. The additional operations introduced by HCC are lightweight. These components require neither an additional optimal-transport solve nor a higher-order tensor construction. We further provide an empirical scaling comparison between KGWOT and the GWOT baseline in Table 4. Across different node scales, KGWOT exhibits a runtime growth trend consistent with GWOT. The measured overhead remains controlled as the graph size increases, confirming that the additional components of KGWOT introduce only a constant-factor cost without changing the asymptotic computational order.

Alignment over Inconsistent Features. While feature perturbation typically leads to performance degradation in conventional graph alignment models, recent GWOT-based studies Xu et al. (2019); Tang et al. (2023a) have developed models exhibiting strong resilience to such perturbations. We evaluate HCC under feature perturbation scenarios in Figure 3 (a), demonstrating that it possesses similar robustness properties. No performance decline is observed across the four datasets, confirming the effectiveness of HCC in handling feature perturbations.

Sensitivity Analysis of Hyperparameters. We present a hyperparameter comprehensive sensitivity study in Figure 4, which analyzes the effect of the learning rate lr in Figure 4 (a), the weights of the entropy term ϵ in Figure 4 (b), the weights of graph distance consistency λ in Figure 4 (c), the scaling factor γ in Figure 4 (d). We find that (1) The model is robust to the lr within a fairly large range, but when the lr exceeds a certain value, there will be a significant performance degradation. (2) When ϵ takes the value in the range of [0.005, 0.1], the model performs stably. However, when ϵ is too large, performance degradation is also observed. At this time, the weight of the entropy term is too large, resulting in a large error in the transport plan. (3) The model is generally robust to λ, but as λ increases, we can observe a certain degree of the growth of the Hit@1 accuracy. (4) A larger γ can lead to better model performance, but when γ is large enough, the model becomes robust to it. In general, our model is robust enough w.r.t these hyperparameters.

5 Conclusion

In this paper, we propose HCC to address the challenges of representation robustness and alignment precision in unsupervised graph alignment. HCC integrates two collaborative components. The GDC module enforces macro-level structural invariance between original and augmented graphs by utilizing inter-graph distance as a regularization term which enables the model to learn noise-resistant representations under the unsupervised learning paradigm. The KGWOT module leverages KKT consistency constraints through a dual step process involving the WAFI and KRIM step to achieve sparse and accurate micro-level node correspondences. Experiments and theoretical analysis demonstrate that HCC promotes structural consistency across different levels of graph alignment and thus contributes to the real-world application of unsupervised graph alignment methods.

AI USE STATEMENT

In this work, we have not used generative AI tools for generating synthetic data sets, helping develop theoretical models or conceptual frameworks, formulating mathematical claims, providing critical ingredients for proving mathematical claims, assisting in the writing of proofs, proposing or refining hypotheses, designing or providing feedback on research methodology or experiments, implementing methods, assisting with translation, cleaning and reformatting dataset, supporting qualitative and thematic data analysis, and interpreting results. We have reviewed all AI-assisted work. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

To facilitate reproducibility, we provide detailed descriptions of the proposed HCC framework in Section 3, with the detailed algorithm summarized in Algorithm 1. The experimental settings, including datasets, model architectures, training configurations, and evaluation metrics, are described in Section 4 and Appendix. We will make the source code publicly available upon acceptance to facilitate independent reproduction of our results and support future research.

REFERENCES

Xavier Aramayo Carrasco, Maksim Nekrashevich, Petr Mokrov, Evgeny Burnaev, and Alexander Korotin. Uncovering challenges of solving the continuous gromov-wasserstein problem. *arXiv preprint arXiv:2303.05978*, 2023.

Liqun Chen, Zhe Gan, Yu Cheng, Linjie Li, Lawrence Carin, and Jingjing Liu. Graph optimal transport for cross-domain alignment. In *International Conference on Machine Learning*, pp. 1542–1553. PMLR, 2020.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013.

Ji Gao, Xiao Huang, and Jundong Li. Unsupervised graph alignment with wasserstein distance discriminator. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, pp. 426–435, 2021.

Ziqi Gao, Chenran Jiang, Jiawen Zhang, Xiaosen Jiang, Lanqing Li, Peilin Zhao, Huanming Yang, Yong Huang, and Jia Li. Hierarchical graph learning for protein–protein interaction. *Nature Communications*, 14(1):1093, 2023.

Mark Heimann, Haoming Shen, Tara Safavi, and Danai Koutra. Regal: Representation learning-based graph alignment. In *Proceedings of the 27th ACM international conference on information and knowledge management*, pp. 117–126, 2018.

Leonid V Kantorovich. On the translocation of masses. *Journal of mathematical sciences*, 133(4), 2006.

Genji Kawakita, Ariel Zeleznikow-Johnston, Naotsugu Tsuchiya, and Masafumi Oizumi. Gromov–wasserstein unsupervised alignment reveals structural correspondences between the color similarity structures of humans and large language models. *Scientific Reports*, 14(1):15917, 2024.

Jure Leskovec and Julian Mcauley. Learning to discover social circles in ego networks. *Advances in neural information processing systems*, 25, 2012.

Jiajin Li, Jianheng Tang, Lemin Kong, Huikang Liu, Jia Li, Anthony Man-Cho So, and Jose Blanchet. Fast and provably convergent algorithms for gromov-wasserstein in graph data. *arXiv preprint arXiv:2205.08115*, 2022.

Qian Li, Shu Guo, Yangyifei Luo, Cheng Ji, Lihong Wang, Jiawei Sheng, and Jianxin Li. Attribute-consistent knowledge graph representation learning for multi-modal entity alignment. In *Proceedings of the ACM web conference 2023*, pp. 2499–2508, 2023.

Xiaoze Liu, Junyang Wu, Tianyi Li, Lu Chen, and Yunjun Gao. Unsupervised entity alignment for temporal knowledge graphs. In *Proceedings of the ACM web conference 2023*, pp. 2528–2538, 2023.

Facundo Mémoli. Gromov–wasserstein distances and the metric approach to object matching. *Foundations of computational mathematics*, 11(4):417–487, 2011.

Facundo Mémoli. The gromov–wasserstein distance: A brief overview. *Axioms*, 3(3):335–341, 2014.

Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.*, pp. 666–704, 1781.

Yu Nesterov. Smooth minimization of non-smooth functions. *Mathematical programming*, 103: 127–152, 2005.

Shichao Pei, Lu Yu, Guoxian Yu, and Xiangliang Zhang. Graph alignment with noisy supervision. In *Proceedings of the ACM web conference 2022*, pp. 1104–1114, 2022.

Gabriel Peyré, Marco Cuturi, and Justin Solomon. Gromov-wasserstein averaging of kernel and distance matrices. In *International conference on machine learning*, pp. 2664–2672. PMLR, 2016.

Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina Eliassi-Rad. Collective classification in network data. *AI magazine*, 29(3):93–93, 2008.

Jianheng Tang, Weiqi Zhang, Jiajin Li, Kangfei Zhao, Fugee Tsung, and Jia Li. Robust attributed graph alignment via joint structure learning and optimal transport. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, pp. 1638–1651. IEEE, 2023a.

Jianheng Tang, Kangfei Zhao, and Jia Li. A fused Gromov-Wasserstein framework for unsupervised knowledge graph entity alignment. In *Findings of the Association for Computational Linguistics: ACL 2023*, pp. 3320–3334. Association for Computational Linguistics, 2023b.

Vayer Titouan, Nicolas Courty, Romain Tavenard, and Rémi Flamary. Optimal transport for structured data with application on graphs. In *International Conference on Machine Learning*, pp. 6275–6284. PMLR, 2019.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

Cédric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2009.

Song Wang, Yushun Dong, Binchi Zhang, Zihan Chen, Xingbo Fu, Yinhan He, Cong Shen, Chuxu Zhang, Nitesh V Chawla, and Jundong Li. Safety in graph machine learning: Threats and safeguards. *arXiv preprint arXiv:2405.11034*, 2024a.

Yejiang Wang, Yuhai Zhao, Daniel Zhengkui Wang, and Ling Li. Galopa: Graph transport learning with optimal plan alignment. *Advances in Neural Information Processing Systems*, 36:9117–9130, 2023.

Yuanyi Wang, Haifeng Sun, Jiabo Wang, Jingyu Wang, Wei Tang, Qi Qi, Shaoling Sun, and Jianxin Liao. Towards semantic consistency: Dirichlet energy driven robust multi-modal entity alignment. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, pp. 3559–3572. IEEE, 2024b.

Zhichun Wang, Qingsong Lv, Xiaohan Lan, and Yu Zhang. Cross-lingual knowledge graph alignment via graph convolutional networks. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, pp. 349–357, 2018.

Felix Wu, Amauri Souza, Tianyi Zhang, Christopher Fifty, Tao Yu, and Kilian Weinberger. Simplifying graph convolutional networks. In *International conference on machine learning*, pp. 6861–6871. Pmlr, 2019.

Hongteng Xu, Dixin Luo, Hongyuan Zha, and Lawrence Carin Duke. Gromov-wasserstein learning for graph matching and node embedding. In *International conference on machine learning*, pp. 6932–6941. PMLR, 2019.

Wenwen Xu and Benjamin CM Fung. Stylelink: User identity linkage across social media with stylometric representations. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 19, pp. 2076–2088, 2025.

Yuguang Yan, Canlin Yang, Yuanlin Chen, Ruichu Cai, and Michael Ng. Hypergraph learning for unsupervised graph alignment via optimal transport. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 21913–21921, 2025.

Si Zhang and Hanghang Tong. Final: Fast attributed network alignment. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1345–1354, 2016.

Si Zhang and Hanghang Tong. Attributed network alignment: Problem definitions and fast solutions. *IEEE Transactions on Knowledge and Data Engineering*, 31(9):1680–1692, 2018.

Si Zhang, Hanghang Tong, Long Jin, Yinglong Xia, and Yunsong Guo. Balancing consistency and disparity in network alignment. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, pp. 2212–2222, 2021.

Marinka Zitnik and Jure Leskovec. Predicting multicellular function through multi-layer tissue networks. *Bioinformatics*, 33(14):i190–i198, 2017.

A Implementation Details

A.1 Sub-Views

$\mathbf{C}_s^{(k)}$ and $\mathbf{C}_t^{(k)}$ are intra-graph measures from different sub-views, i.e., edge-view, node-view and subgraph-view, where K is the number of sub-views.

- **Edge-view:** The edge-view $C_s^{(1)}$and $C_t^{(1)}$ encode the first-order topology only. if two nodes v_i and v_j share an edge, $C_{s,ij}^{(1)}$ and $C_{t,ij}^{(1)}$are close, otherwise they are infinitely far. No learnable parameters are introduced; the raw adjacency is adopted for the construction of the edge-view as $C_s^{(1)} = \mathcal{E}_s, C_t^{(1)} = \mathcal{E}_t$.
- **Node-view:** When the attribute vectors of two nodes are alike, we treat them as structurally close even if no edge exists. This view is complementary to the edge-view and is robust to missing/perturbed edges. A linear similarity matrix is built with inner products as $C_s^{(2)} = \mathcal{Z}_s^T \mathcal{Z}_s, C_t^{(2)} = \mathcal{Z}_t^T \mathcal{Z}_t$.
- **Subgraph-view:** Capture k-hop neighborhood patterns without learning any weight matrices. We diffuse the attributes along the graph for k steps and then compute the smoothed vectors for q-th subgraph-view, $2 < q \leq K$,

$$\boldsymbol{C}_s^{(q)} = \hat{\mathcal{E}}_s^{q-2} \mathcal{Z}_s (\hat{\mathcal{E}}_s^{q-2} \mathcal{Z}_s)^T, \boldsymbol{C}_t^{(q)} = \hat{\mathcal{E}}_t^{q-2} \mathcal{Z}_t (\hat{\mathcal{E}}_t^{q-2} \mathcal{Z}_t)^T, \tag{16}$$

where $\hat{\mathcal{E}}_s = D^{-1/2}(\mathcal{E}_s + I)D^{-1/2}$ and $\hat{\mathcal{E}}_t = D^{-1/2}(\mathcal{E}_t + I)D^{-1/2}$. D_s and D_t are the degree matrix of $\mathcal{E}_s$ and $\mathcal{E}_t$ respectively.

A.2 DATASETS.

To evaluate the performance of HCC in terms of both robustness and accuracy, we conduct experiments on four synthetic datasets under various difficulty settings. Specifically, on each dataset, we introduce varying degrees of structural perturbation noise to evaluate the model's robustness against structural disturbances. Since existing methods have already achieved strong robustness to feature perturbations, this paper focuses on analyzing the impact of structural perturbations on model performance.

- Cora Sen et al. (2008): A citation network of academic papers in the field of machine learning, where nodes correspond to scientific publications and edges denote citation relationships. Each paper is assigned to one of seven predefined classes.
- Citeseer Sen et al. (2008): Another citation network, similar to Cora, but larger and sparser. It consists of computer science publications as nodes, categorized into six research topics.
- PPI Zitnik & Leskovec (2017): A protein–protein interaction network comprising human proteins as nodes. Edges indicating biological interactions.
- Facebook Leskovec & Mcauley (2012): A subgraph extracted from the Facebook social network. Nodes represent users and edges represent connections. Node features are constructed by user profile information.

Additionally, we assess the real-world graph alignment capability of HCC using the following datasets:

- Douban Online-Offline Zhang & Tong (2016): A social network collected from Douban, a popular Chinese platform for user-generated reviews and recommendations of books, movies, and music.
- ACM-DBLP Zhang & Tong (2018): A bibliographic network dataset constructed from two major computer science bibliography sources, ACM and DBLP. It includes authorship information. The task is to identify authors from both sources that refer to the same individual.

A.3 BASELINES.

Baselines. We benchmark HCC against a diverse suite of seven unsupervised graph alignment methods, each representing a distinct technical approach.

- REGAL Heimann et al. (2018) generates node embeddings via structural feature engineering and performs alignment in the embedding space using a similarity-based matching.
- GCNAlign Wang et al. (2018) leverages Graph Convolutional Networks (GCNs) to learn node representations by propagating and aggregating neighborhood information.
- GATAlign Veličković et al. (2017) builds upon GCNs by incorporating Graph Attention Networks (GATs), which assigns different weights to neighbors.
- WAlign Gao et al. (2021) is an unsupervised method that iteratively refines node embeddings and the alignment matrix.
- GWD Xu et al. (2019) frames graph alignment as a Gromov-Wasserstein discrepancy problem. It leverages pairwise intra-graph distances for matching, making it well-suited for purely structural alignment where node features are unavailable.
- FusedGW Titouan et al. (2019) extends the Gromov-Wasserstein framework by fusing it with standard optimal transport.
- SLOTAlign Tang et al. (2023a) is a recent robust method that employs structure-aware optimal transport, potentially incorporating higher-order structural information or using robust loss functions to handle noisy graph data effectively.
- HLOT Yan et al. (2025) is a recent baseline that introduces hypergraph as an additional sub-view.

We will open-source the code of the proposed HCC model once accepted.

B PROOF FOR PROPOSITION

Proposition 1. [The iteration method of WAFI] For the problem of Eq. (9), we leverage the fixed-point method and obtain a new iteration formula as follows with $\mathcal{K}(f_j) = \sum_{k=1}^{M} \exp\left(\frac{f_k - \bar{L}_{kj}^{(T)}}{\epsilon}\right)$:

$$f_i^{(t+1)} = \epsilon \left[\log(\mu_{s,i}) - \log \sum_{j=1}^{N} \left[\mu_{t,j} \exp\left(-\frac{\bar{L}_{ij}^{(T)}}{\epsilon}\right) \mathcal{K}^{-1}\left(f_j^{(t)}\right) \right] \right]. \tag{17}$$

Proof. For notational convenience, we first introduce the partition functions:

$$\mathcal{H}_j\left(\boldsymbol{f}^{(\ell)}\right) = \sum_{i=1}^{M} \exp\left(\frac{f_i^{(\ell)} - \bar{L}_{ij}^{(T)}}{\epsilon}\right). \tag{18}$$

Therefore we can obtain the following results:

$$\exp\left(\frac{f_i^{(\ell)}}{\epsilon}\right) \sum_{j=1}^{N} \left[\frac{b_j \exp\left(-\frac{\bar{L}_{ij}^{(T)}}{\epsilon}\right)}{\sum_{k=1}^{M} \exp\left(\frac{f_k^{(\ell)} - \bar{L}_{kj}^{(T)}}{\epsilon}\right)} \right] = a_i \exp\left(\frac{f_i^{(\ell)} - f_i^{(\ell+1)}}{\epsilon}\right). \tag{19}$$

Next, expressing the kernel evaluated at the updated stage then we have the results:

$$\begin{cases} \exp\left(\dfrac{f_i^{(\ell)} - \bar{L}_{ij}^{(T)}}{\epsilon}\right) = \mathcal{H}_j\left(\boldsymbol{f}^{(\ell)}\right) \dfrac{\pi_{ij}^{(\ell)}}{b_j} \\ \exp\left(\dfrac{f_i^{(\ell+1)} - \bar{L}_{ij}^{(T)}}{\epsilon}\right) = \exp\left(\dfrac{f_i^{(\ell)} - \bar{L}_{ij}^{(T)}}{\epsilon}\right) \dfrac{a_i}{\sum_{k=1}^{N} \pi_{ik}^{(\ell)}}. \end{cases} \tag{20}$$

Finally we can compute the difference in the dual objective as:

$$\begin{aligned} \Delta &= \mathcal{J}_{FL}\left(\boldsymbol{f}^{(\ell)}\right) - \mathcal{J}_{FL}\left(\boldsymbol{f}^{(\ell+1)}\right) \\ &= \sum_{j=1}^{N} \epsilon b_j \log \left[\frac{\sum_{i=1}^{M} \exp\left(\frac{f_i^{(\ell)} - C_{ij}}{\epsilon}\right)}{\sum_{i=1}^{M} \exp\left(\frac{f_i^{(\ell+1)} - C_{ij}}{\epsilon}\right)} \right] - \sum_{i=1}^{M} \left(f_i^{(\ell)} - f_i^{(\ell+1)}\right) a_i \\ &= \epsilon \sum_{j=1}^{N} b_j \log\left(\frac{b_j}{\rho_j^{(\ell)}}\right) + \epsilon \sum_{i=1}^{M} a_i \log\left(\frac{a_i}{\gamma_i^{(\ell)}}\right) \\ &= \epsilon D_{\mathrm{KL}}\left(\boldsymbol{b} \| \boldsymbol{\rho}^{(\ell)}\right) + \epsilon D_{\mathrm{KL}}\left(\boldsymbol{a} \| \boldsymbol{\gamma}^{(\ell)}\right) \\ &\geq 0, \end{aligned} \tag{21}$$

where we denote $\boldsymbol{\rho}^{(\ell)} = \sum_{i=1}^{M} \frac{a_i \pi_{ij}^{(\ell)}}{\sum_{k=1}^{N} \pi_{ik}^{(\ell)}} = \sum_{i=1}^{M} \frac{a_i \pi_{ij}^{(\ell)}}{\gamma_i^{(\ell)}}$ and $\boldsymbol{\gamma}^{(\ell)} = \sum_{j=1}^{N} \pi_{ij}^{(\ell)}$. Therefore we can conclude that $\mathcal{J}_{FL}\left(\boldsymbol{f}^{(\ell+1)}\right) \leq \mathcal{J}_{FL}\left(\boldsymbol{f}^{(\ell)}\right)$ which indicates the convergence. □