CLUHCS: Dual-View Contrastive Learning Enabled Unsupervised Heterogeneous Community Search with Meta-Path Behavior Modeling

Xiaoqin Xie[1], Bin Zhao[1], Mingzhu Chang[1], Shuai Han[1*], Wu Yang[1]

[1]College of Computer Science and Technology, Harbin Engineering University, Harbin, China
{xiexiaoqin, binzhao, changmingzhu, hshuai, yangwu}@hrbeu.edu.cn

Abstract

Existing community search methods heavily rely on labeled data or predefined structures, thus fail to capture obscure and dynamic community boundaries in open-world heterogeneous networks, leading to poor adaptability. They also ignore modeling behavioral patterns, resulting in poor search performance. To solve the above issues, this work formally defines the unsupervised behavior-driven community search problem for heterogeneous graphs and designs dual-view Contrastive Learning-based Unsupervised framework for Heterogeneous graph Community Search (CLUHCS). CLUHCS designs a relation view to encode local community cohesion and a meta-path view to capture global behavior semantics. By using PathSim averaging strategy to generate positive samples and self-supervised signals, we can completely eliminate label dependency. Then, contrastive training is leveraged to automatically learn community representations and solve the open community boundary ambiguity challenge. Furthermore, by capturing behavior patterns, the meta-path behavior modeling flexibly characterizes the formation mechanism of heterogeneous communities. Experiments on three datasets verify the effectiveness and efficiency of CLUHCS. CLUHCS significantly improves F1-score by 52.7% over the supervised baseline FCS-HGNN and by 41.5% over the unsupervised method TransZero.

Code and data — https://github.com/Iinfrg/CLUHCS.

Introduction

Heterogeneous graphs, with their richer semantics and complex relationships compared to homogeneous graphs, contain more valuable community structures. Community search on such graphs, which aims to find a cohesive subgraph containing query nodes, is a critical task with applications in recommendation systems, fraud detection, gene function prediction (Zhao et al. 2022; Li et al. 2022a), and knowledge graph link prediction (Li et al. 2022b).

Traditional community search methods use predefined models to represent community structures, eliminating the need for labeled data and enabling generalization to unknown communities. However, these models impose strict topological constraints, leading to "structural rigidity" (Xie et al. 2024), while actual communities may not match such theoretical conditions. For example, k-core methods (Sozio and Gionis 2010; Fang et al. 2020) mandate that each node's degree must be at least k, which is often unsuitable for boundary nodes in real-world communities. To address this issue, recent research introduces learning-based search methods that alleviate dependency on rigid predefined models. For example, TransZero (Wang et al. 2024a) implements self-supervised community search in homogeneous graphs via pre-trained graph Transformers.

However, learning-based community search on heterogeneous graphs faces two fundamental limitations in open-world scenarios. The first is excessive reliance on annotated data. High-quality community labels are scarce since the community boundaries are too obscure and dynamic to capture, while existing methods exacerbate this through rigid supervision. For instance, FCS-HGNN (Chen et al. 2024) requires full community labels for training, and MK-HGNN's self-training strategy (Li et al. 2024) still demands initial annotations. Such dependency severely impairs generalization to unseen real-world communities. Therefore, we present a contrastive learning framework to eliminate the dependency on labeled data. The second is neglecting of behavioral patterns, leading to poor community search performance. Current approaches focus narrowly on local structural and attribute features via GNNs, such as QDGNN (Jiang et al. 2022), but inherent defects like over-smoothing prevent them from capturing global behavioral semantics. Namely, they fail to model the consistent behavior patterns that the community members follow to form their community. Recent community search methods (Jiang et al. 2022; Zhou et al. 2023) enumerate all the possible meta-paths without differentiating the behavior semantics that different meta-paths carry. They find indistinguishable meta-path neighbors for further searching of a specific community. This leads to free-rider effect that the nodes irrelevant to query nodes could be included in the target community. Furthermore, a deeper analysis of how behavioral patterns themselves influence and shape high-level network structures (e.g., communities) is still lacking. Therefore, we aim to design a meta-path Transformer and PathSim-based approach to model the community behavioral patterns. The core objective is to analyze how different types of behavioral patterns drive and influence community formation in heterogeneous networks.

*Corresponding author

To overcome the above challenges, we propose a novel problem, called Unsupervised Behavior-driven Heterogeneous Community Search (UBHCS) problem. Then, we propose the first unsupervised framework that eliminates label dependency through contrastive self-supervision and discovers behavior-aware communities via meta-path modeling.

Specifically, as shown in Figure 1, we construct a dual-view contrastive learning framework comprising four modules. First, the type-aware feature mapping module projects features of different node types into a shared space. Then, the relation-aware GNN module dynamically considers the contribution of various relations in the heterogeneous graph from the relation view. It aggregates local neighborhood information to produce nodes' local feature embeddings. Furthermore, the meta-path transformer module expands the receptive field via K-hop neighbors and generates token sequences from the meta-path view. After passing through a graph transformer, it applies token-level and semantic-level attention to produce nodes' global feature embeddings. Later, the contrastive learner measures node community feature similarity by using the meta-path averaged PathSim metric, which models the nodes' behavior. Meanwhile, we propose a positive sampling strategy. By intra-view and inter-view learning, this strategy can help the model automatically capture community features and identify the dynamic community boundaries.

In this paper, the first contribution is proposing the first unsupervised contrastive learning framework for heterogeneous community search, which eliminates labeled data dependency while adapting to dynamic community boundaries. Another contribution is that we design a meta-path average PathSim-based positive sampling strategy that captures the community's behavioral patterns through multi-path semantic fusion. While meta-paths have been used for feature extraction, we propose a PathSim-Averaging strategy that repurposes them to generate self-supervised signal for contrastive learning. This paradigm shift differs from passive feature propagation in GNNs and supervised meta-path weighting schemes. Finally, we empirically validate CLUHCS against state-of-the-art baselines on 3 datasets. Experimental results demonstrate that CLUHCS outperforms baseline methods across all metrics.

Related Work

Traditional Community Search. Traditional community search identifies cohesive subgraphs using predefined structural models such as k-core and k-truss (Sozio and Gionis 2010; Fang et al. 2020). While methods like D-core (Fang et al. 2018), EquiTruss (Akbas and Zhao 2017), and CSRTI (Xie et al. 2024) efficiently find dense subgraphs in homogeneous graphs, they face two main limitations: structural rigidity, which excludes communities that violate strict topologies, and the assumption of homogeneity, overlooking node and edge heterogeneity. Extensions incorporating meta-paths (Jiang et al. 2022; Liu et al. 2024; Fang et al. 2020) attempt to address heterogeneity but neglect consistent behavioral semantics. They also fail to capture behavioral coherence and lack flexibility, as they rely on static or manually tuned metrics. Our CLUHCS overcomes these issues by replacing rigid, handcrafted metrics with contrastive behavior similarity learning, enabling adaptive, semantic-aware community search.

Learning-based Community Search. Learning-based methods address structural rigidity, falling into two main categories. Homogeneous graph methods explore data-driven search, with early supervised models like ICS-GNN (Gao et al. 2021) and IACS (Fang et al. 2024) suffering from annotation dependence and retraining cost. Semi-supervised methods like COCLEP (Li et al. 2023), and the self-supervised models like TransZero (Wang et al. 2024a) and CSFormer (Wang et al. 2024c) still rely on labels, limiting their applicability in annotation-scarce environments. COCLEP and TransZero are all for homogeneous graph while heterogeneous graph methods remain under-explored. FCS-HGNN (Chen et al. 2024) fuses multi-relational signals without meta-path constraints, while MK (Li et al. 2024) combines attributes and structure via meta-paths but limits fixed community sizes. ALICE (Wang et al. 2024b) extracts candidate subgraph by GNN-based ConNet. All these methods depend on supervision, constraining their generalization. Our CLUHCS framework removes supervision requirements and preserves structural adaptability with a significant advancement in community search.

Heterogeneous Graph Representation Learning. Heterogeneous graph representation learning aims to embed multi-typed nodes and edges into unified spaces. Existing methods fall into shallow and deep categories. Shallow models like metapath2vec (Dong, Chawla, and Swami 2017) and HIN2Vec (Fu, Lee, and Lei 2017) use meta-paths with random walks or proximity constraints, limited in modeling complex interactions. Deep models such as HGT (Hu et al. 2020) and R-GCNs (Schlichtkrull et al. 2018) directly capture heterogeneity but risk semantic confusion, while meta-path-based techniques like HAN (Wang et al. 2019), Se-HGNN (Yang et al. 2023), and BPHGNN (Fu et al. 2023) aggregate semantically relevant neighbors. HGMS (Wang et al. 2025) leverages multi-view self-expression for contrastive learning. However, these supervised methods face three limitations, (1) limited behavioral analysis; (2) heavy reliance on labels or pretext tasks without applicability in open-world scenarios; (3) task-method misalignment, i.e., general objectives like node classification or link prediction do not effectively optimize community-coherent representations. Our CLUHCS method uses meta-path behavior modeling and self-supervised dual-view learning, offering a flexible and effective community search solution.

The Method

Problem Definition

Let $HG = (\mathcal{V}, \mathcal{E}, \mathcal{T}, \mathcal{R}, \mathcal{X}, \varphi, \psi)$ be the given heterogeneous graph, where $\mathcal{V}$ and $\mathcal{E}$ denote the node set and edge set, $\mathcal{X}$ represents node attributes, $\mathcal{T}$ and $\mathcal{R}$ are node-type set and edge-type set satisfying $|\mathcal{T}| + |\mathcal{R}| > 2$. We define node and edge type mapping functions, namely $\varphi : \mathcal{V} \to \mathcal{T}$ assigns types to vertices, while $\psi : \mathcal{E} \to \mathcal{R}$ categorizes edges.

A meta-path ϕ in a heterogeneous graph describes the association between nodes or basic behavior patterns, formally defined as $\phi : \mathcal{T}_1 \xrightarrow{R_1} \mathcal{T}_2 \xrightarrow{R_2} \cdots \xrightarrow{R_l} \mathcal{T}_{l+1}$, signifying a composite relation $\mathcal{R} = \mathcal{R}_1 \circ \mathcal{R}_2 \circ \cdots \circ \mathcal{R}_l$ between node types $\mathcal{T}_1$ and $\mathcal{T}_{l+1}$, where $\circ$ denotes the relation composition operator and l represents the meta-path length.

Definition 1 (Meta-Path Average PathSim.) Given a node v, meta-path set Φ, and neighbor $u \in \cup_{\phi\in\Phi} N^1_\phi(v)$, where $\mathcal{N}^k_\phi(v)$ is the set of k-hop neighbors of v under a meta-path ϕ, the meta-path average PathSim $PS(v,u)$ (Sun et al. 2011) serves as self-supervised signal for contrastive learning:

$PS(v,u) = \frac{1}{|\Phi|}\sum_{\phi\in\Phi}\frac{2\times|\{p_{v\to u}|p_{v\to u}\vdash\phi\}|}{|\{p_{v\to v}|p_{v\to v}\vdash\phi\}|+|\{p_{u\to u}|p_{u\to u}\vdash\phi\}|}$

where $p_{v\to u} \vdash \phi$ means $p_{v\to u}$ is the path instance of meta-path ϕ. Our meta-path averaging strategy can generate self-supervised signals without label dependency.

Definition 2 (Target Type Community.) Given a set of query nodes V_q of type T_q in HG, the target type node set $\mathcal{V}_{T_q} \subseteq \mathcal{V}$ comprises all nodes sharing type T_q. Nodes of auxiliary types $\mathcal{T}_{aid} = \mathcal{T}\backslash\{T_q\}$ serve solely as structural and semantic context during community search. Crucially, the output community only contains nodes from $\mathcal{V}_{T_q}$, while auxiliary nodes are excluded from output communities.

Definition 3 (Single Node Community Score.) Given a HG, a query node set $V_q \subseteq \mathcal{V}_{T_q}$, and learned embeddings $Z = \{\mathbf{z}_v \in \mathbb{R}^d | v \in \mathcal{V}_{T_q}\}$, the single node community score for node $v \in \mathcal{V}_{T_q}$ is(Wang et al. 2024a):

$\mathrm{ss}_v = \frac{1}{|V_q|}\sum_{q_i\in V_q}\cos(\mathbf{z}_v, \mathbf{z}_{q_i}) = \frac{1}{|V_q|}\sum_{q_i\in V_q}\frac{\mathbf{z}_v^\top \mathbf{z}_{q_i}}{\|\mathbf{z}_v\|_2\|\mathbf{z}_{q_i}\|_2}$

This score quantifies the behavioral consistency between v and the query nodes, where $\|\cdot\|_2$ denotes the l_2-norm. Higher ss_v means stronger alignment with the community structure.

Definition 4 (UBHCS Problem.) Given a heterogeneous graph HG=($\mathcal{V}$,$\mathcal{E}$,$\mathcal{T}$,$\mathcal{R}$,$\mathcal{X}$,φ, ψ) and a query node set $V_q \subseteq \mathcal{V}_{T_q}$ of target type T_q, the goal of UBHCS is to find the connected community $C_q \subseteq \mathcal{V}_{T_q}$ containing V_q that maximizes the **Expected Score Gain (ESG)** (Wang et al. 2024a),

argmax$ESG(C_q)$=argmax$_{C_q}\frac{1}{|C_q|^\omega}\left(\sum_{v\in C_q} ss_v - \mu|C_q|\right)$

where $\mu = \frac{\sum_{u\in V} ss_u}{|V|}$ is the global score mean, ss_v is the aggregate similarity between the embeddings of v and query nodes, both of which encode their behavior semantics. $\omega \in [0,1]$ controls community cohesion. Crucially, the goal is reached with no community supervision, using only graph structural and node behavioral signals. We chose ESG because it evaluates communities using embeddings, aligning well with our unsupervised objective.

Framework of Unsupervised Heterogeneous Community Search

We propose CLUHCS (Contrastive Learning-based Unsupervised Heterogeneous Community Search) framework, the first unsupervised framework for community search in heterogeneous graphs. As shown in Figure 1, CLUHCS comprises four key components, i.e., Type-Aware Feature Mapping Module, Relation-Aware GNN Module, Meta-Path Transformer Module, and Contrastive Learner.

Particularly, unlike generic contrastive learning, our dual-view architecture is purpose-built for community search. The relation view preserves local community cohesion and the meta-path view captures global behavioral semantics. This joint design directly tackles the open boundary ambiguity challenge. CLUHCS eliminates label dependency via self-supervised contrastive training and captures behavioral patterns by PathSim metric. The trained model $\mathcal{M}$ generates node embeddings $Z = \mathcal{M}(HG)$. For query V_q, the community $C_q = F(HG, V_q, \mathcal{M})$ is obtained by maximizing ESG.

Type-Aware Feature Mapping. This module aims to resolve feature heterogeneity across types and preserve type semantics in embedding space. For node $v \in \mathcal{V}$ of type $\varphi(v) = t$, we project features to a type-consistent space $h_v = \sigma(W_t \cdot x_v + b_t)$, where W_t is a type-specific projection matrix, x_v is initial feature embedding, h_v is the projected embedding in sharing space.

Relation-Aware GNN Module. This module takes node embeddings after feature mapping as input and outputs the local feature embedding for the target type nodes. This module aggregates neighborhood information while weighting edge relations dynamically from the relation view as follows: $h^l_v = \sigma\left(\sum_{u\in\mathcal{N}_v}\alpha^l_{vu}W^l h^{l-1}_u + W^l_1 h^{l-1}_v\right)$, where α^l_{vu} = softmax(MLP $(h^{l-1}_v \| h^{l-1}_u \| e_{vu})$) adaptively weights relation r_{vu}'s contribution, and e_{vu} is the feature vector of the edge (v,u). We stack all target type nodes embedding and get a feature matrix $Z^{rv} \in \mathbf{R}^{n\times d}$, n is the number of the target type nodes, d is the feature dimension.

Meta-Path Transformer Module. This module models the semantic and structural information. It captures both the long-range dependencies along specific meta-path and the diverse semantic information across different meta-paths. As illustrated in Figure 1, the module operates through four key stages: (1) Meta-path aware Token Generation, (2) Meta-path specific Transformer Encoding, (3) Token-level Attention, and (4) Semantic-level Attention.

(1) Meta-Path Aware Token Generation: To enable Transformer to handle graphs, a given node's and its neighbors' representations are converted into token sequences. For node v, meta-path ϕ, and maximum hop count K, we generate a sequence of tokens representing the node and its neighbors at different hops under ϕ. Specifically, the k-hop token ($k = 0, 1, 2, \ldots, K$) for all target nodes is obtained by aggregating features from their k-hop neighbors $\mathcal{N}^k_\phi(v)$ (where k=0 corresponds to the node itself). This aggregation is efficiently computed for the graph by using the normalized adjacency matrix based on the meta-path based adjacency:

$$X^k_\phi = (\hat{A}_\phi)^k H, \quad (k = 0, 1, 2, \ldots, K)$$

where ($\hat{A}_\phi = D^{-1/2}A_\phi D^{-1/2}$) is the normalized adjacency matrix for meta-path ϕ. A_ϕ is the original adjacency matrix induced by ϕ, D is the corresponding degree matrix, $H \in \mathbb{R}^{n\times d}$ is the initial node feature matrix, and X^k_ϕ represents the k-hop token matrix for the target node type. The token sequence for node v under meta-path ϕ is then $S_{\phi,v} = [X^0_{\phi,v}; X^1_{\phi,v}; \ldots; X^K_{\phi,v}]$. This strategy effectively integrates both node features and multi-hop behavioral semantics specific to the meta-path ϕ. For the meta-path set Φ, we obtain the collection of token sequences $S_\Phi = \{S_\phi | \forall\phi \in \Phi\}$,

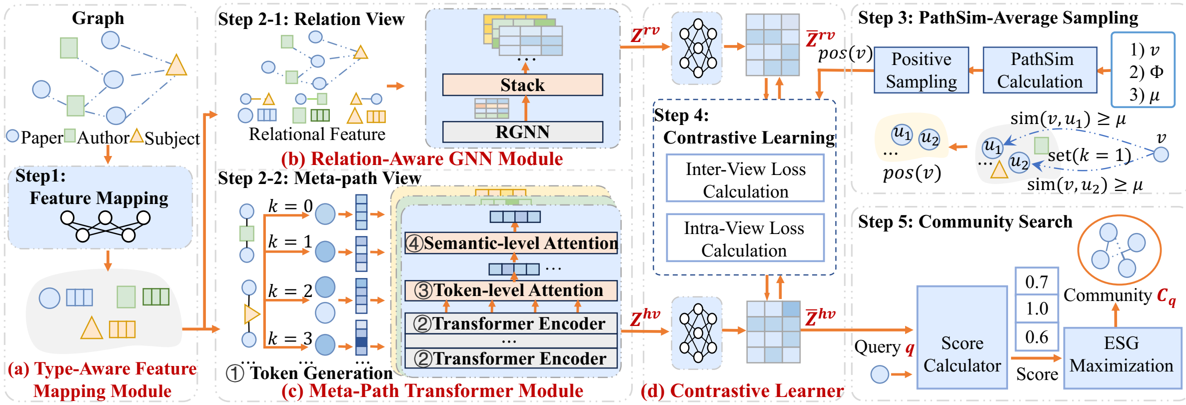


Figure 1: Illustration of the **CLUHCS** framework. **(1) Step 1. Type-Aware Feature Mapping:** Projects heterogeneous node features into type-consistent latent space via W_t. **(2) Step 2. Dual-View Encoding:** Relation View aggregates ***local*** neighborhoods with dynamic edge weighting, while Meta-Path View captures ***global*** behaviors via multi-hop path attention. **(3) Step 3. PathSim-Average Sampling:** Generates self-supervised signals by fusing multi-path similarities. **(4) Step 4. Contrastive Learner:** Aligns ***local / global*** views by $\mathcal{L}$. **(5) Step 5. Community Search:** Outputs C_q by maximizing ESG.

where $S_\phi = [X_\phi^0; X_\phi^1; \ldots; X_\phi^K]$.

(2) Meta-path Specific Transformer Encoding: Each token sequence $S_{\phi,v}$ is independently fed into a dedicated Transformer encoder with L_{tm} layers. The Transformer encoder captures complex interactions and dependencies between tokens at different hops within the same meta-path sequence. The encoder output for $S_{\phi,v}$ is an enhanced token representation matrix $Z_{\phi,v} \in \mathbb{R}^{(K+1)\times d}$, and each row corresponds to a contextualized representation of a hop token.

(3) Token-level Attention: Not all hop tokens contribute equally to the final node representation under a specific meta-path. To dynamically learn the importance of each hop relative to the node itself (the 0-hop token), we apply a token-level attention mechanism (Chen et al. 2023):

$$\alpha_{\phi,v}^k = \frac{\exp\big((Z_{\phi,v}^0 \| Z_{\phi,v}^k) W_{\phi,ta}^T\big)}{\sum_{k'=1}^K \exp\big((Z_{\phi,v}^0 \| Z_{\phi,v}^{k'}) W_{\phi,ta}^T\big)}$$

where $\alpha_{\phi,v}^k$ represents the attention weight of the k-hop token for node v under meta-path ϕ, $Z_{\phi,v}^k$ denotes the k-hop token of node v under meta-path ϕ, and $W_{\phi,ta} \in \mathbb{R}^{1\times 2d}$ is the learnable weight vector, and $\|$ denotes concatenation. The node representation under meta-path ϕ is then fused as:

$$Z_{\phi,v} = Z_{\phi,v}^0 + \textstyle\sum_{k=1}^K \alpha_{\phi,v}^k Z_{\phi,v}^k$$

This step produces a meta-path specific node embedding $Z_{\phi,v} \in \mathbb{R}^d$ that integrates information from relevant hops.

(4) Semantic-level Attention: Different meta-paths represent distinct behavioral patterns and semantics. To learn the importance of a meta-path $\phi \in \Phi$ for final node representation, we employ a semantic-level attention mechanism:

$$\beta_{\phi,v} = \frac{\exp(\delta_\phi \tanh(W_\phi z_{\phi,v}))}{\sum_{\phi' \in \Phi} \exp\big(\delta_{\phi'} \tanh\big(W_{\phi'} z_{\phi',v}\big)\big)}$$

where $\delta_\phi \in \mathbb{R}^{d_s}$ and $W_\phi \in \mathbb{R}^{d_s \times d}$ are learnable parameters. The global meta-path view embedding Z_v^{hv} for node v is obtained by aggregating the meta-path embeddings weighted by their learned importance, $z_v^{hv} = \sum_{\phi\in\Phi} \beta_{\phi,v} z_{\phi,v}$. Collecting embeddings for all target nodes yields the meta-path view representation matrix $Z^{hv} \in \mathbb{R}^{n\times d}$.

Contrastive Learner with PathSim-Averaged Sampling. To enable unsupervised community feature learning, we design a dual-view contrastive learner that generates self-supervision signals through meta-path behavioral consensus. Unlike supervised methods (Li et al. 2023) that require explicit community annotations, our key insight recognizes that nodes sharing high behavioral similarity under meta-paths likely belong to the same community. It motivates the design of meta-path averaged PathSim Sampling strategy. Namely, $\forall v \in V$, we compute its multi-path fused similarity with 1-hop neighbors $u \in \cup_{\phi\in\Phi} \mathcal{N}_\phi^1(v)$ as Definition 1. Specifically, the average PathSim $s_{avg}(v,u)$ between v and its neighbor u is computed over all meta-paths $\phi \in \Phi$. Each neighbor u satisfying $s_{avg}(v,u) \geq \mu$ (a predefined threshold) is added into v's positive sample set $pos(v)$. Existing methods compute meta-path similarities as node features, while our method reformulates PathSim as positive sample selectors. The strategy effectively identifies nodes with similar community contexts or behavioral pattern, simulating community annotations crucial for self-supervised training. The key novelty of our method is using PathSim's distribution to quantify community behavior for sampling, because its symmetric, type-specific similarity can natively capture the behavioral consistency our method requires.

To align the dual-view representations for effective comparison, we project both embeddings into a shared contrastive space by using a learnable transformation:

$$\bar{z}_v^{rv} = \sigma_t\left(W_{proj} z_v^{rv} + b_{proj}\right),\ \bar{z}_v^{hv} = \sigma_t(W_{proj} z_v^{hv} + b_{proj})$$

Then, we yield the projected embeddings $\bar{z}^{rv}, \bar{z}^{hv} \in \mathbb{R}^{n\times d}$.

Subsequently, **intra-view learning** enforces behavioral consistency between v and its positives $pos(v)$ within each view while promoting separation from negatives. For a view $\gamma \in \{rv, hv\}$, the loss for v is:

$$\mathcal{L}_{\text{intra}}^\gamma(v) = -\frac{1}{|pos(v)|} \sum_{u\in pos(v)} \log \frac{\exp(\cos(\bar{z}_v^\gamma, \bar{z}_u^\gamma)/\tau)}{\sum_{t\in V} \exp(\cos(\bar{z}_v^\gamma, \bar{z}_t^\gamma)/\tau)}$$

where τ is a temperature parameter.

Crucially, **inter-view learning** captures cross-view consistency by treating the projection $\bar{z}_v^{rv}$ and $\bar{z}_u^{hv}$ (where $u \in pos(v)$) as a positive pair, while contrasting against negatives from other nodes. The loss for node v is:

$$\mathcal{L}_{\text{inter}}(v) = \frac{1}{2|pos(v)|} \sum_{u \in pos(v)} \left[I\left(\bar{z}_v^{rv}, \bar{z}_u^{hv}\right) + I\left(\bar{z}_v^{hv}, \bar{z}_u^{rv}\right) \right]$$

where the contrastive term $I(a,b)$ for any pair is defined as:

$$I(a,b) = -\log \frac{\exp(\cos(a,b)/\tau)}{\sum_{t \in V} \exp\left(\cos(a, \bar{Z}_t^{\gamma_b})/\tau\right)}$$

with γ_b denoting the view of b (i.e., rv or hv).

Then, the total self-supervised loss $\mathcal{L}_s$ combines intra-view and inter-view losses across all nodes:

$$\mathcal{L}_s = \frac{1}{|V|} \sum_{v \in V} \left(\frac{1}{2} \left(\mathcal{L}_{\text{intra}}^{rv}(v) + \mathcal{L}_{\text{intra}}^{hv}(v) \right) + \lambda \mathcal{L}_{\text{inter}}(v) \right)$$

where λ balances the inter-view importance.

Furthermore, we leverage the inherent alignment between views by treating different view projections of the same node v as a natural positive pair, defined by an unsupervised loss:

$$\mathcal{L}_u = \frac{1}{2|V|} \sum_{v \in V} \left[I\left(\bar{z}_v^{rv}, \bar{z}_v^{hv}\right) + I\left(\bar{z}_v^{hv}, \bar{z}_v^{rv}\right) \right]$$

To sum up, the final contrastive objective combines these components to jointly optimize behavioral consistency and view alignment for community coherence, denoted as,

$$\mathcal{L} = \mathcal{L}_s + \alpha \frac{1}{|V|} \sum_{v \in V} \mathcal{L}_u(v)$$

where α controls the alignment weight. This loss jointly optimizes community-aware node similarity within and across views while preserving their intrinsic correspondence.

As a result, the contrastive learner eliminates label dependency while capturing community structures through meta-path behavioral patterns, which is the key innovation for enabling unsupervised community search.

Model Training and Community Search Algorithm

Training Dynamics and Unsupervised Optimization. CLUHCS undergoes end-to-end optimization through dual-view contrastive learning, eliminating dependency on community labels by using self-supervised signals derived from meta-path behavioral patterns. During training, heterogeneous features are first projected into a type-consistent latent space via parametric mapping. The relation-aware GNN then encodes local structural proximity, while the meta-path Transformer captures global behavioral semantics by hierarchical attention over multi-hop neighbors. *Our core innovation* lies in the contrastive learner's use of meta-path averaged PathSim similarity to generate positive samples, effectively simulating community signals without supervision. The unified loss $\mathcal{L}$ optimizes all parameters via gradient descent, exhibiting two key characteristics. First, synchronized parameter updates ensure feature consistency across views. Second, the convergence stabilizes when PathSim-based samples have high similarity in the contrastive space. This alignment confirms robust community feature extraction. After training and optimization, the output is model $\mathcal{M}$.

Behavior-Aware Community Search. Upon convergence, the trained model $\mathcal{M}$ generates behavior-aware embedding $Z = \mathcal{M}(HG)$ that encodes intrinsic community structures. For query V_q, the search algorithm computes community scores s_v by Definition 3, quantifying behavioral similarity to the query set. After initializing C_q as V_q, the solution expands iteratively by greedy ESG maximization:

$$v^* = \arg\max_{v \in \mathcal{N}(C_q)} \Delta\text{ESG}(v) = \arg\max_{v \in \mathcal{N}(C_q)} (\text{ESG}(C_q \cup v) - \text{ESG}(C_q))$$

where $\mathcal{N}(C_q)$ denotes the set of unvisited nodes that are connected to any node in the set C_q through any meta-path.

Our algorithm advances prior works from three aspects. First, we overcome limitations of homogeneous methods by adapting to heterogeneous behavioral patterns via meta-path fused embeddings. Second, by ω-adjusted penalization without predefined size constraints, we realize open-world granularity control. Third, we use PathSim-based positive sampling strategy to learn self-supervised signals and inject community features into node embeddings. It ensures effective behaviorally-cohesive community search.

Complexity Analysis

The computational complexity of CLUHCS is analyzed across training and search phases. Let $|\mathcal{V}|$ and $|\mathcal{E}|$ be the node and edge counts. d is the embedding dimension, $n = |\mathcal{V}_{T_q}|$ is the target node size, and $|\Phi|$ is the meta-path number. During training, each epoch has four primary operations. Feature mapping costs $O(|\mathcal{V}|d^2)$ time due to type-specific projections. The relation-aware GNN module costs $O(L_g(|\mathcal{V}| + |\mathcal{E}|)d^2)$ time (L_g is the GNN layer number) for neighborhood aggregation, while the meta-path Transformer demands $O(n|\Phi|L_{tm}Kd^2)$ time (L_{tm} is the Transformer layer number) for behavioral pattern encoding. The contrastive learner costs $O(n^2\bar{M}d)$ time where $\bar{M}$ denotes average meta-path neighbors. Thus, the training time complexity over t epochs is $O(t \cdot ((|\mathcal{V}| + |\mathcal{E}| + n|\Phi|)d^2 + n^2\bar{M}d))$.

For community search, there are three sequential steps. Embedding inference first generates behavior-aware representations in $O(|(V)| + n|\Phi|d^2)$ time. Community scoring then computes node-query similarities in $O(|V_q|nd)$ time, although this becomes negligible as $|V_q| \ll n$. Finally, greedy expansion with heap-based optimization performs community growth in $O(n^2 \log n)$ time. The overall search complexity thus reduces to $O((|\mathcal{V}| + n|\Phi|)d^2 + n^2 \log n)$.

Experiments

Experimental Setup

Datasets. We use three commonly-used real-life heterogeneous datasets(Table1). ACM (Yu et al. 2024) models academic publications with three node types: paper, author, and subject. Here, paper serve as target node type. DBLP (Fu et al. 2020) captures computer science publications through four node types: authors, papers, conferences, and terms, targeting authors. IMDB (Fu et al. 2020) shows film network with movie, actor, and director nodes, targeting movies.

Dataset	Node number	Edge number	Meta-path	Average degree
ACM	paper(P):4019 author(A):7167 subject(S):60	P-A:13407 P-S:4019	PAP PSP	3.1
DBLP	author(A):4057 paper(P):14328 conference(C):20 term(T):7723	P-A:19645 P-T:85810 P-C:14329	APA APCPA APTPA	9.2
IMDB	movie(M):4278 director(D):2081 actor(A):5257	M-D:4278 M-A:12828	MAM MDM	2.9

Table 1: Dataset Summary.

Compared Methods. We compare our approach against three state-of-the-art methods representing key paradigm. **(1) FCS-HGNN**(Chen et al. 2024) is a learning-based method using multi-view extraction and semantic attention. **(2) WC-index**(Sun et al. 2022) serves as a traditional homogeneous baseline with a two-tier index and refinement. **(3) TransZero**(Wang et al. 2024a) is an unsupervised method employing pre-trained graph Transformers with offline learning and only similarity search. WC-index and TransZero should be adapted to heterogeneous setting via meta-path projection. The baselines cover key methodological categories (learning-based, index-driven, unsupervised) and handle both homogeneous and heterogeneous settings.

Metrics. We mainly adopt Precision, Recall, F1-score, and Jaccard similarity that are commonly used.

Implementation Details. PyTorch 2.1.0 with DGL 2.4.0 on an Ubuntu 22.04 system is adopted, with an NVIDIA RTX 4090 GPU and Intel Xeon Gold 6430 CPU. The GNN contains 2 layers while the Transformer uses 4 layers with 8 multi-head attention units. The embedding dimension is fixed at 64 across all datasets. Key hyperparameters including hop count $K \in [1,7]$ (step 1), ESG granularity control $\omega \in [0,1]$ (step 0.1), and PathSim sampling threshold $\mu \in [0.1,1]$ (step 0.1), are tuned via grid search, with optimal values selected per dataset. All experiments execute 40 times and compute the average. To assess the robustness of our model, we initialize five independent models with varying random seeds. The observed standard deviations analysis results demonstrate exceptional model stability.

Experiment Design. We design four experiments to evaluate the effectiveness and efficiency of our CLUHCS model.

1. Performance Comparison Experiments. We evaluate the performance of CLUHCS against three baselines. For supervised FCS-HGNN baseline, we implement three query strategies (the split of train/validation/test is 3:1:5). (1) **Inductive (FCS-HGNN-I)**: Communities are divided roughly in 1:1 ratio. Train/val queries come from training set while test queries are from unseen test set. This setup evaluates generalization. (2) **Transduction (FCS-HGNN-T)**: All queries are sampled from full community pool (full exposure). (3) **Hybrid (FCS-HGNN-H)**: Communities are split roughly in 1:1 ratio. Train/val queries come from training set while test queries are from all communities. This setup simulates partial-knowledge deployment. WC-index and TransZero are evaluated by using meta-path projections to relevant node types without requiring specialized training split.

2. Ablation Experiments. To quantify component contributions, CLUHCS-A replaces selective positive sampling with all meta-path neighbors, and CLUHCS-Z eliminates positive sampling entirely. We further ablate the dual-view framework via CLUHCS-MP (no meta-path view) and CLUHCS-Rel (no relation view), isolating the complementary value of structural semantics versus direct interactions. A search-layer ablation is also discussed.

3. Parameter Sensitivity Analysis. CLUHCS's hyperparameter sensitivity of four critical parameters is evaluated: the hop number K that governs neighbor aggregation scope, the sampling threshold μ controlling positive sample selection, and the expected score parameter ω balancing community cohesion. It provides crucial insights into model robustness and guides configuration.

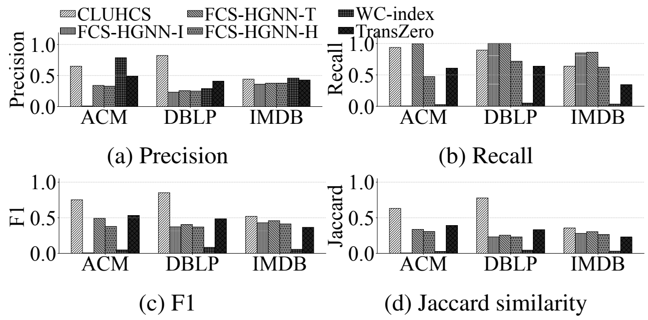


Figure 2: Performance comparison of different methods.

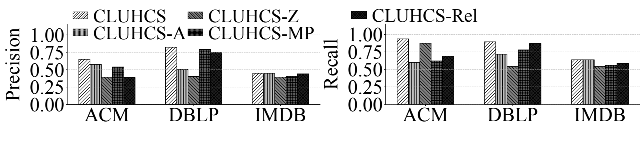
4. Efficiency Evaluation Experiment. We rigorously evaluate time efficiency against competitors, measuring both training time and online search latency. Since WC-index lacks trainable components, only its search efficiency is assessed, demonstrating CLUHCS's deployment viability.

Experimental Results

Performance Comparison. In Figure 2, CLUHCS consistently outperforms all baselines on 3 datasets in F1-score and Jaccard similarity, with particularly substantial improvements on ACM and DBLP. For example, CLUHCS significantly improves F1-score by 52.7% over the unsupervised FCS-HGNN-T and by 41.5% over supervised TransZero. It is due to two key designs of CLUHCS. One is meta-path filtering strategy which excludes irrelevant nodes and boosts precision. The other is contrastive learning which identifies boundary nodes and maintains competitive recall. It proves our framework advances unsupervised community search instead of merely applying existing ML tools.

In addition, despite achieving higher recall (reaching recall=1.0 on DBLP), FCS-HGNN-I and FCS-HGNN-T exhibit significantly lower precision due to their tendency to return overly large communities. While these communities cover most target nodes, they include excessive irrelevant nodes, resulting in substantially lower F1 scores than CLUHCS, which achieves an optimal precision-recall balance. FCS-HGNN-T's performance is higher than FCS-HGNN-I/H and the reason would be its inherent limitation of dependence on known community structures and poor generalization. TransZero's underperformance validates the necessity of explicit heterogeneous information modeling. These results demonstrate CLUHCS's efficacy as an unsupervised solution for high-accuracy community search.

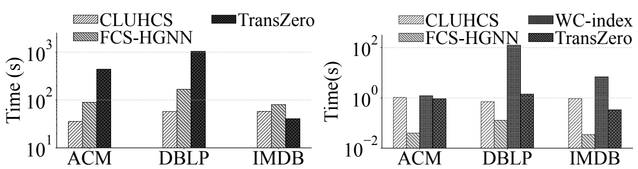
Ablation Experiment. As shown in Figure 3, CLUHCS outperforms all ablated variants on ACM and DBLP. On IMDB, it matches CLUHCS-A while surpassing other variants. The performance degradation of CLUHCS-Z across all datasets and metrics confirms that meta-path neighbors provide essential self-supervision for community characteristics. CLUHCS-A's inferior results on ACM/DBLP demon-



(a) Precision (b) Recall

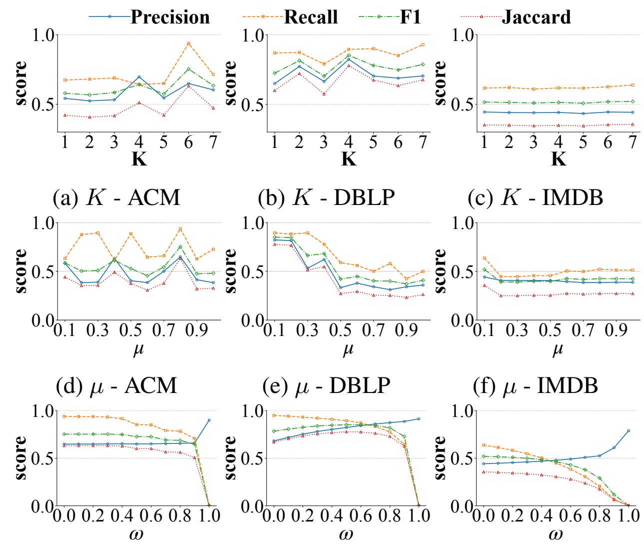
(c) F1 (d) Jaccard similarity

Figure 3: Ablation results of different methods.



(a) Training time (b) Search time

Figure 4: Efficiency evaluation of different methods.



(a) K - ACM (b) K - DBLP (c) K - IMDB

(d) μ - ACM (e) μ - DBLP (f) μ - IMDB

(g) ω - ACM (h) ω - DBLP (i) ω - IMDB

Figure 5: Performance on different parameter values.

strate that unfiltered meta-path neighbors introduce harmful noise, validating our PathSim-based positive sampling selection. Notably, CLUHCS-A achieves comparable performance to CLUHCS on IMDB, since its meta-path neighbors inherently exhibit that low noise and uniform community characteristics require no selective filtering.

CLUHCS-MP's consistent underperformance validates the necessity of meta-path view, while CLUHCS-Rel's deficiencies confirm the importance of relational view. These results collectively demonstrate the complementary value of both views. In addition, in a search-layer ablation with fixed embeddings, ESG proved optimal by balancing quality and scale. Conversely, clustering and top-k search methods improved with larger granularity (fewer clusters or higher k).

Efficiency Analysis. In Figure 4, CLUHCS achieves superior training efficiency. CLUHCS completes training within one minute on all datasets, while FCS-HGNN and TransZero consume double time on ACM and DBLP. Though longer than TransZero on IMDB, CLUHCS remains competitive with consistent sub-minute performance.

All learning-based methods achieve sub-second search latency. FCS-HGNN exhibits the fastest performance by eliminating explicit similarity computations, while CLUHCS and TransZero show comparable efficiency despite requiring pairwise node-query similarity measurements. Conversely, traditional WC-index incurs high search latency, particularly on dense DBLP networks. It highlights the practical superiority of learning-based methods for real-time applications.

Parameter Sensitivity Analysis. CLUHCS exhibits robust yet dataset-dependent performance across key parameters. For neighborhood hops(K) in Figure 5(a)-(c), ACM and DBLP display significant sensitivity due to their strong reliance on meta-path view features, aligning with ablation results that meta-path removal caused substantial F1/Jaccard decrease. In contrast, IMDB shows minimal fluctuation as its meta-path view contributes marginally. Optimal values are K=6 (ACM), K=4 (DBLP), and K=7 (IMDB).

Sampling threshold μ controls neighbor selection rigor, where higher μ enforces stricter similarity. As shown in Figure 5(d)-(f), the analysis reveals that ACM uniquely achieves peak performance at μ=0.8, confirming that tight similarity thresholds yield highly accurate communities. Conversely, DBLP and IMDB perform optimally at μ=0.1, reflecting their tolerance to noisier neighbors. This performance variation indicates that optimal μ adapts to dataset characteristics.

Parameter ω controls community cohesiveness, where higher values yield smaller, tighter communities. In Figure 5(g)-(i), the performance establishes a precision-recall tradeoff, i.e., increasing ω elevates precision but reduces recall. On ACM and IMDB, F1 and Jaccard decline with ω, indicating ground-truth communities favor loose structures. Conversely, DBLP exhibits an inverse U-curve peaking at ω=0.5. It is because insufficient cohesion admits irrelevant nodes (low precision). It proves moderate cohesion can optimize performance for fine-grained communities.

Conclusion

In this paper, we propose CLUHCS, a contrastive learning enabled unsupervised heterogeneous community search method. It resolves obscure community boundaries and label dependency by discovering behavior-driven communities in open-world networks. This capability stems from joint modeling of local structural semantics (relation-aware GNN) and global behavioral dynamics (meta-path Transformers), where PathSim-based positive sampling generates self-supervised behavioral signals for robust community characterization. Experiments validate superior performance over state-of-the-art baselines.

Acknowledgements

This research was sponsored by the NSFC-Regional Joint Fund Key Program under Grant U22A2036, NSFC-Key Program of Enterprise Joint Fund under Grant U21B2019, the National Natural Science Foundation of China under Grant 62272127 and 62572144, and the Natural Science Foundation Joint Guidance Project of Heilongjiang under Grant LH2024F036.

References

Akbas, E.; and Zhao, P. 2017. Truss-based community search: a truss-equivalence based indexing approach. *Proceedings of the VLDB Endowment*, 10(11): 1298–1309.

Chen, G.; Guo, F.; Wang, Y.; Liu, Y.; Yu, P.; Shen, H.; and Cheng, X. 2024. Fcs-hgnn: Flexible multi-type community search in heterogeneous information networks. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, 207–217.

Chen, J.; Gao, K.; Li, G.; and He, K. 2023. NAGphormer: A Tokenized Graph Transformer for Node Classification in Large Graphs. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*.

Dong, Y.; Chawla, N. V.; and Swami, A. 2017. metapath2vec: Scalable representation learning for heterogeneous networks. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 135–144.

Fang, S.; Zhao, K.; Rong, Y.; Li, Z.; and Yu, J. X. 2024. Inductive attributed community search: to learn communities across graphs. *Proceedings of the VLDB Endowment*, 17(10): 2576–2589.

Fang, Y.; Wang, Z.; Cheng, R.; Wang, H.; and Hu, J. 2018. Effective and efficient community search over large directed graphs. *IEEE Transactions on Knowledge and Data Engineering*, 31(11): 2093–2107.

Fang, Y.; Yang, Y.; Zhang, W.; Lin, X.; and Cao, X. 2020. Effective and efficient community search over large heterogeneous information networks. *Proceedings of the VLDB Endowment*, 13(6): 854–867.

Fu, C.; Zheng, G.; Huang, C.; Yu, Y.; and Dong, J. 2023. Multiplex Heterogeneous Graph Neural Network with Behavior Pattern Modeling. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2023, Long Beach, CA, USA, August 6-10, 2023*, 482–494.

Fu, T.-y.; Lee, W.-C.; and Lei, Z. 2017. Hin2vec: Explore meta-paths in heterogeneous information networks for representation learning. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 1797–1806.

Fu, X.; Zhang, J.; Meng, Z.; and King, I. 2020. Magnn: Metapath aggregated graph neural network for heterogeneous graph embedding. In *Proceedings of the web conference 2020*, 2331–2341.

Gao, J.; Chen, J.; Li, Z.; and Zhang, J. 2021. ICS-GNN: lightweight interactive community search via graph neural network. *Proceedings of the VLDB Endowment*, 14(6): 1006–1018.

Hu, Z.; Dong, Y.; Wang, K.; and Sun, Y. 2020. Heterogeneous graph transformer. In *Proceedings of the web conference 2020*, 2704–2710.

Jiang, Y.; Fang, Y.; Ma, C.; Cao, X.; and Li, C. 2022. Effective community search over large star-schema heterogeneous information networks. *Proceedings of the VLDB Endowment*, 15(11): 2307–2320.

Li, L.; Luo, S.; Zhao, Y.; Shan, C.; Wang, Z.; and Qin, L. 2023. Coclep: Contrastive learning-based semi-supervised community search. In *2023 IEEE 39th International Conference on Data Engineering*, 2483–2495.

Li, M.; Cai, X.; Li, L.; Xu, S.; and Ji, H. 2022a. Heterogeneous graph attention network for drug-target interaction prediction. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 1166–1176.

Li, Y.; Chen, X.; Zhao, Y.; Shan, W.; Wang, Z.; Yang, G.; and Wang, G. 2024. Self-training gnn-based community search in large attributed heterogeneous information networks. In *2024 IEEE 40th International Conference on Data Engineering*, 2765–2778.

Li, Z.; Zhao, Y.; Zhang, Y.; and Zhang, Z. 2022b. Multi-relational graph attention networks for knowledge graph completion. *Knowledge-Based Systems*, 251: 109262.

Liu, Y.; Guo, F.; Xu, B.; Bao, P.; Shen, H.; and Cheng, X. 2024. SACH: Significant-Attributed Community Search in Heterogeneous Information Networks. In *2024 IEEE 40th International Conference on Data Engineering*, 3283–3296.

Schlichtkrull, M.; Kipf, T. N.; Bloem, P.; Van Den Berg, R.; Titov, I.; and Welling, M. 2018. Modeling relational data with graph convolutional networks. In *The semantic web: 15th international conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, proceedings 15*, 593–607. Springer.

Sozio, M.; and Gionis, A. 2010. The community-search problem and how to plan a successful cocktail party. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, 939–948.

Sun, L.; Huang, X.; Li, R.; Choi, B.; and Xu, J. 2022. Index-Based Intimate-Core Community Search in Large Weighted Graphs. *IEEE Transactions on Knowledge and Data Engineering*, 34(9): 4313–4327.

Sun, Y.; Han, J.; Yan, X.; Yu, P. S.; and Wu, T. 2011. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. *Proceedings of the VLDB Endowment*, 4(11): 992–1003.

Wang, H.; Shi, C.; Xu, C.; Yan, S.; Xie, R.; and Tang, P. 2025. Enhancing Homophily in Heterogeneous Graph Contrastive Learning via Connection Strength and Multi-view Self-Expression. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2025, Padua, Italy, July 13-18, 2025*, 709–718.

Wang, J.; Wang, K.; Lin, X.; Zhang, W.; and Zhang, Y. 2024a. Efficient Unsupervised Community Search with Pre-trained Graph Transformer. *Proc. VLDB Endow.*, 17(9): 2227–2240.

Wang, J.; Wang, K.; Lin, X.; Zhang, W.; and Zhang, Y. 2024b. Neural Attributed Community Search at Billion Scale. *Proceedings of the ACM on Management of Data*, 1(4): 1–25.

Wang, X.; Ji, H.; Shi, C.; Wang, B.; Ye, Y.; Cui, P.; and Yu, P. S. 2019. Heterogeneous graph attention network. In *The world wide web conference*, 2022–2032.

Wang, Y.; Gou, X.; Xu, X.; Geng, Y.; Ke, X.; Wu, T.; Yu, Z.; Chen, R.; and Wu, X. 2024c. Scalable community search over large-scale graphs based on graph transformer. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1680–1690.

Xie, X.; Liu, S.; Zhang, J.; Han, S.; Wang, W.; and Yang, W. 2024. Efficient Community Search Based on Relaxed k-Truss Index. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1691–1700.

Yang, X.; Yan, M.; Pan, S.; Ye, X.; and Fan, D. 2023. Simple and efficient heterogeneous graph neural network. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, 10816–10824.

Yu, J.; Ge, Q.; Li, X.; and Zhou, A. 2024. Heterogeneous graph contrastive learning with meta-path contexts and adaptively weighted negative samples. *IEEE Transactions on Knowledge and Data Engineering*, 36(10): 5181–5193.

Zhao, B.-W.; Hu, L.; You, Z.-H.; Wang, L.; and Su, X.-R. 2022. HINGRL: predicting drug–disease associations with graph representation learning on heterogeneous information networks. *Briefings in bioinformatics*, 23(1): bbab515.

Zhou, Y.; Fang, Y.; Luo, W.; and Ye, Y. 2023. Influential community search over large heterogeneous information networks. *Proceedings of the VLDB Endowment*, 16(8): 2047–2060.