**Edelweiss Applied Science and Technology**
*ISSN: 2576-8484*
*Vol. 10, No. 8, 36-47*
*2026*
*Publisher: Learning Gate*
*DOI: 10.55214/2576-8484.v10i8.13428*


# Hybrid-HGT: A unified deep learning framework for community detection in heterogeneous attributed graphs

Hamza Haddad[1]*, Hicham Attariuas[2], Ali Younes[3]
[1,2,3]Department of Informatics, Faculty of Sciences, University Abdelmalik Saadi, Morocco; hamza.haddad@uae.ac.ma (H.H.) hattarius@uae.ac.ma (H.A.), ayounes@uae.ac.ma (A.Y.).

**Abstract:** Identifying communities in heterogeneous attributed graphs remains a significant challenge for networks consisting of multi-membership nodes and multi-relations, attribute-rich information, and very few labels. Here, we present a combined HGT-based framework incorporating contrastive representation learning and deep clustering with multi-round training via pseudo-labels. This enables us to jointly utilize both structural connections and semantic similarities in order to yield more informative node embeddings and discover interesting community structures. Experiments on DBLP and IMDB benchmark datasets demonstrate that the proposed framework is able to achieve competitive performance, verifying its effectiveness for community detection in real-world heterogeneous networks. In addition, the iterative pseudo-labeling mechanism forms a powerful learning strategy that progressively increases supervision by conferring confident predictions on currently unlabeled nodes, while the contrastive and clustering objectives facilitate representation discrimination and community cohesion, enabling our model to better deal with label scarcity, heterogeneous dependencies, and overlapping semantics in practical, complex, attributed networks.
***Keywords:*** *Clustering objectives, Community detection, Contrastive representation learning, Heterogeneous graph, Transformers.*

## 1. Introduction

Finding communities is a significant problem for making sense of the organization of complex networks, which can be many large groups of nodes that behave similarly, share characteristics, or interact with each other [1]. It is used in several areas, like citation networks, social networks, and multimedia recommendation systems. This really means that many existing community detection methods work well only on homogeneous graphs, while most real-world networks are heterogeneous, with multi-type nodes, context, and additional semantic attributes [2]. Such heterogeneously structured networks bring in new modeling complexities that traditional approaches are unable to accommodate [1, 3, 4]. Academic collaboration networks, social media platforms, and movie recommender systems provide three well-known examples of this. Graph neural networks (GNNs) are one of the recent approaches to graph representation learning, which improve training with neighborhood aggregation and feature propagation mechanisms [5, 6]. However, many of the existing GNN-based methods either convert heterogeneous graphs to homogeneous graphs or impoverish semantic information and node attributes [7]. Furthermore, a lot of the state-of-the-art models are heavily reliant on supervised learning and therefore need vast quantities of labeled data that are not usually available for real-world use cases [8-10]. A solution to these constraints is proposed in this work, which incorporates the Heterogeneous Graph Transformer (HGT) architecture [11] into a hybrid framework for community detection on heterogeneous attributed graphs. It incorporates type-aware attention, contrastive representation learning, clustering-aware optimization, and multi-round pseudo-labeling to learn semantically localized node embeddings robustly with weak supervision that is highly dependent on the labeled information.


**History:** Received: 11 May 2026; Revised: 6 July 2026; Accepted: 10 July 2026; Published: 7 August 2026
* Correspondence: hamza.haddad@uae.ac.ma

The pseudo-labeling method iteratively expands the supervision set with reliable predictions to enhance the model's generalization performance in weakly supervised scenarios. We conduct experiments of the proposed framework on two heterogeneous benchmark datasets, DBLP and IMDB, that have distinct structural and semantic characteristics. Experimental results demonstrate that the proposed method achieves consistently better clustering results than several competitive baseline approaches, including MAGNN [12] and DMGI [13], based on Normalized Mutual Information (NMI) and Adjusted Rand Index (ARI). In summary, this work presents a holistic hybrid community detection framework that unifies supervised learning, contrastive learning, clustering refinement, and iterative pseudo-label propagation, achieving strong empirical performance on real-world heterogeneous graph datasets.

## 2. Related STUDY

Community detection in complex networks has been a very prominent area of research, with state-of-the-art developments incorporating the ideas of GNNs, semantic learning, and contrastive objectives. In this Section, we overview several important directions related to our work, such as classic and deep community detection approaches, heterogeneous graph representation learning, and semi-supervised frameworks. Table 1. Summary comparison of discussed models.

**Table 1.**
Comparative Capabilities of Community Detection Methods on Heterogeneous Graphs.

| Method | Graph Type | Uses Node Attributes | Heterogeneous Support | Contrastive Learning | Clustering Objective | Pseudo-Labeling | Target Task |
|---|---|---|---|---|---|---|---|
| HAN | Heterogeneous | ✓ | ✓ | ✗ | ✗ | ✗ | Node classification |
| MAGNN | Heterogeneous | ✓ | ✓ | ✗ | ✗ | ✗ | Node classification |
| HAESF | Heterogeneous | ✓ | ✓ | ✗ | ✓ | ✗ | Community detection |
| DMGI | Heterogeneous | ✓ | ✓ | ✓ | ✗ | ✗ | Representation learning |
| BP-GCN | Heterogeneous | ✓ | ✓ | ✗ | ✓ | ✗ | Community detection |
| Our(Hybrid-HGT) | Heterogeneous | ✓ | ✓ | ✓ | ✓ | ✓ | Community detection |

### *2.1. Community Detection in Graphs*

Established community detection approaches like modularity optimization, spectral clustering, and label propagation have been very successful on homogeneous graphs, where both the types and structures of nodes and edges are consistent. Yet, these strategies are not sufficient for heterogeneous graphs with various types of nodes and relationships. The rise of heterogeneous networks with greater structural and semantic complexity has stimulated the design of deep learning methods to learn node embeddings in a way that community structures can be represented more clearly [14, 15]. To our knowledge, HAESF [16] is the first to utilize a semantic factorization framework with selective attention mechanisms for community detection in large-scale heterogeneous graphs. BP-GCN [17] also adopted a self-attention mechanism over type-aware relations to better characterize heterogeneous interactions and improve clustering strength. Despite being particularly tailored to heterogeneous community detection, the existing approaches still possess lackluster support for contrastive representation learning and semi-supervised pseudo-labeling strategies, as outlined in Table 1.

### *2.2. Representation Learning on Heterogeneous Attributed Graphs*

HetGNNs have been introduced to address these limitations, with promising progress achieved by transforming graphs with different node types and relationship kinds into heterogeneous sample representations, allowing us to learn richer and more informative representations. One of the strongest methods, HAN (Heterogeneous Attention Network) [18], proposed a metapath-based attention mechanism to capture semantic relationships in heterogeneous structures. MAGNN [12] expanded the idea by utilizing information from intermediate nodes along metapaths to preserve more explicit semantic dependencies for the model. These methods obtain strong performance on the node classification task, but their formulations are not designed for clustering or community detection. On the other hand, DMGI [13] uses a multi-view contrastive learning approach to train robust unsupervised node representations by maximizing consistency between different semantic views. Although DMGI performs well in representation learning, it does not directly optimize clustering objectives and lacks mechanisms that utilize label information or pseudo-supervised learning to facilitate community formation.

### *2.3. Contrastive and Clustering-Aware Learning*

Recently, contrastive learning has successfully become a widely used unsupervised representation learning method by pulling embeddings for similar samples closer together while pushing away those that are dissimilar [19]. Numerous famous techniques, such as GRACE [20] and DGI [21], have produced excellent results in homogeneous graph contexts. However, these methods are not fully capable of producing distinct and semantically meaningful communities when generalized to heterogeneous graphs, as is practiced by DMGI. As shown in Table 1, few heterogeneous graph models simultaneously integrate contrastive learning and clustering-oriented optimization in a unified framework. To bridge this gap, we propose a Hybrid-HGT framework that jointly conducts contrastive representation learning and deep clustering objectives, explicitly regularizing a group of similar nodes to be arranged compactly, yet distanced from other unmatched community members.

### *2.4. Semi-Supervised Learning with Pseudo-Labeling*

Since many real-world graphs contain very few labeled nodes, interest in researching pseudo-labeling/self-training-based strategies for graph representation learning has increased. Though approaches like MERIT [22] have demonstrated excellent performance in semi-supervised settings on homogeneous networks, heterogeneous graphs as inputs are still a very new topic. Also note that some up-to-date heterogeneous graph models (e.g., HAN [18], MAGNN [12], BP-GCN [17]) do not combine self-training or iterative pseudo-labeling mechanisms, as pointed out in Table 1. To overcome this limitation, we devise a multi-round pseudo-labeling approach to progressively grow the supervision set with highly confident predictions in our proposed Hybrid-HGT framework. It iteratively propagates the original information from nodes to edges, which reinforces the structure-aware clustering and semantic consistency of communities, thereby boosting model performance with limited labeled data.

## 3. Methodology

In this work, the problem we aim to solve is how to learn node embeddings to simultaneously preserve structural dependencies and semantic coherence for detecting communities in heterogeneous attributed graphs. To this end, we design a unified framework integrating heterogeneous graph transformers, contrastive representation learning, deep clustering, and pseudo-labeling under one model description. This framework jointly learns the graph topology and semantic information and endeavors to generate more discriminative embeddings for community discovery in complex heterogeneous networks. The next section introduces the proposed pipeline components and how they interact for enhanced clustering performance.

### *3.1. Problem Definition*

We formulate the problem on a heterogeneous attributed graph G=(V, E, X, T), where V denotes the set of nodes with different types, E is the typed edge set [23], X refers to node attributes or feature representations, and T represents node and edge categories. Problem: Tag target nodes (e.g., papers in

DBLP, movies in IMDB) to meaningful community clusters that reflect latent semantic regularities or functional similarity, given a partially labeled bipartite graph. Solving this problem raises two key challenges for the case of heterogeneous graphs, i.e., learning rich representations for both nodes and relations while accurately identifying community structures in large environments where labeled data are scarce.

### *3.2. Overview of the Framework*

The framework integrates supervised node classification, contrastive learning, and finally clustering-aware optimization under a single learning framework. It involves supervised learning on a few labeled nodes, contrastive learning using semantic consistency between node representations, and clustering-aware learning with the aim of refining community assignments by optimizing deep embeddings. We optimize these objectives jointly in a Heterogeneous Graph Transformer (HGT) framework that accurately models complex relational structures and scores type-specific interactions on heterogeneous graphs. In contrast to traditional approaches, where clustering or pseudo-labeling are seen as independent post-processing components, our framework seamlessly connects each of them into a fully end-to-end training model. Such optimization allows representation learning, clustering refinement, and supervision expansion to interact in a unified manner for iterative improvements in embedding quality and the coherence of community structures.

### *3.3. Heterogeneous Graph Transformer (HGT)*

The HGT module is the core building block of our framework, through which it can learn type-aware node embeddings through attention-based heterogeneous neighborhood aggregation of relation-specific information. We first project each node feature vector into a shared latent space, where original node features are mapped through a learnable matrix of the corresponding type (Equation (1)). These transformed representations are further passed through the HGT layers, as in Equation (2), to capture both topological relationships and cross-node/structural dependencies for each semantic type of node across all types (metapaths). The embeddings that follow include more information about the structure and latent semantics of the graph. In the case of node classification, the obtained embeddings are finally connected to a linear classifier followed by a softmax layer, as in Equation (3), to jointly learn the classifier parameters during training for predicting the final node classes.

$$h(v) = W_{\phi(\mathrm{v})} x_v \tag{1}$$

$$z_v = HGT(h_v, N(v), E) \tag{2}$$

$$\hat{y}_v = softmax(W_c z_v + b_c) \tag{3}$$

### *3.4. Contrastive Pretraining*

We incorporate a contrastive learning objective into our framework based on the InfoNCE formulation to boost robustness and semantic quality of node representations. Instead of explicit graph augmentation relaxations such as GraphMask, the model generates another view of the embedding via a stochastic shuffle that produces implicit positive and negative pairs. The cosine similarity function defined in Eq. (4) is used to measure the similarity of embedding representations, which enables the framework to evaluate semantic consistency with more than one representation of embedding views. From these similarity scores, the contrastive objective defined by (5) encourages embeddings of correlated nodes to stay close in the latent space and pushes away representations of uncorrelated ones, where $\tau$ controls a temperature scaling factor [24]. In order to further stabilize training and provide consistency among the embedding views, a symmetric contrastive formulation is adopted, as illustrated in Eq. (6). This bidirectional objective alleviates representation collapse and enhances the separability and coherence of identified communities.

.

$$sim(z_i^{(1)}, z_j^{(2)}) = \frac{z_i^{(1)} . z_j^{(2)}}{\left\|z_i^{(1)}\right\| \left\|z_j^{(2)}\right\|} \quad (4)$$

$$\mathrm{L}_{con}^{(1)} = -\sum_{i=1}^{N} log \frac{\exp(sim(z_i^{(1)}, z_j^{(2)})/\tau)}{\sum_{j=1}^{N} \exp(sim(z_i^{(1)}, z_j^{(2)})/\tau)} \quad (5)$$

$$L_{con} = \frac{1}{2}(\mathrm{L}_{con}^{(1)} + \mathrm{L}_{con}^{(2)}) \quad (6)$$

*3.5. Clustering-Aware Learning*

To explicitly promote the learning of cohesive communities, we embed a deep clustering loss in our framework, which is based on the Student's t-distribution. We compute soft assignments from node embeddings to cluster centroids with the following formulation, which is defined in (7), where μj denotes the centroid of cluster j, which gives the assignment amount and indicates how similar the nodes need to be in order to match these clusters better. We design a sharpened target distribution to boost clustering confidence and cluster purity, which is expressed mathematically in Eq. (8), where higher-confidence assignments are enhanced, while lower assignments are decreased in impact on the overall decision being made. We then minimize the Kullback–Leibler divergence objective in equation (9) between the predicted soft assignments and their corresponding target distribution. To ameliorate such instabilities in the clustering behavior early on in training, we only introduce the clustering objective after a warm-up stage where embeddings can learn initial useful structure- and semantic-aware representations before enforcing stronger compactness and separation of clusters.

$$q_{ij} = \frac{\left(1+\left\|z_i - \mu_j\right\|^2\right)^{-1}}{\sum_{j'}\left(1+\left\|z_i - \mu_{j'}\right\|^2\right)^{-1}} \quad (7)$$

$$p_{ij} = \frac{q_{ij}^2 / \sum_i q_{ij}}{\sum_{j'}\left(q_{ij'}^2 / \sum_i q_{ij'}\right)} \quad (8)$$

$$L_{cluster} = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (9)$$

*3.6. Multi-Round Pseudo-Labeling*

To address the scarcity of labeled data, the proposed framework incorporates a multi-round pseudo-labeling strategy that is fully integrated into the training process. At fixed intervals, specifically every 50 epochs, the model predicts labels for previously unlabeled nodes, and these predictions are transformed into probability distributions using the softmax operation defined in Equation (10). To reduce the impact of noisy supervision, pseudo-labels are assigned only to nodes whose prediction confidence satisfies the threshold condition described in Equation (11), ensuring that only highly reliable predictions are added to the training set. The final pseudo-label assignment is then determined according to Equation (12). Unlike conventional semi-supervised approaches that perform pseudo-labeling only once, the proposed framework continuously updates and refines pseudo-labels over multiple training rounds, allowing the embedding space and supervision set to evolve simultaneously. This iterative mechanism creates a feedback process in which improved node representations lead to more accurate predictions, while the expanded supervision further strengthens representation learning. The pseudo-labeling procedure progressively converges when no additional unlabeled nodes satisfy the confidence threshold defined in Equation (11) and validation performance becomes stable. Compared with existing semi-supervised methods, the proposed approach differs by integrating pseudo-labeling directly into the end-to-end optimization pipeline, combining dynamic multi-round supervision expansion with both contrastive learning and clustering-aware objectives to achieve more robust representation learning and improved community detection performance in heterogeneous graphs.

$$p_i = softmax(\hat{y}_i) \tag{10}$$

$$max(p_i) > \delta \tag{11}$$

$$\tilde{y}_i = \arg max(p_i) \tag{12}$$

*3.7. Total Loss and Optimization*

The complete optimization includes supervised classification, contrastive learning, and clustering refinement in one training process. The overall achievable objective function consists of supervised classification loss, contrastive representation learning optimization guidance, and a clustering-aware regularization term, as expressed in Equation (13), enabling the model to optimize semantic representation learning as well as community structure refinement. The supervised part is computed by the cross-entropy loss formulation defined in Equation (14), where ground truth supervision comes from labeled nodes during training. The hyperparameters α and β, which determine the importance of the contrastive objective and clustering objectives, respectively, are both set to 0.1 in the experiments to maintain a balanced optimization process. All components of the framework can be updated jointly through backpropagation, as model training is end-to-end with the Adam optimizer. This integrated optimization method preserves structural adjacency, semantic relevance, and community information in the learned embeddings at once to enhance the clustering performance as well as generalization ability.

$$L = L_{sup} + \alpha L_{con} + \beta L_{cluster} \tag{13}$$

$$L_{sup} = \sum_{i \epsilon y_L} \sum_{c=1}^{C} y_{ic} \log(\hat{y}_{ic}) \tag{14}$$

## 4. Experimental Setup

The proposed Hybrid-HGT framework is evaluated on two widely adopted heterogeneous network benchmarks, namely DBLP and IMDB. The DBLP dataset contains heterogeneous entities such as papers, authors, venues, and terms, with papers grouped into four research categories: Database, Data Mining, Artificial Intelligence, and Information Retrieval. The IMDB dataset includes movies, actors, and directors, with each movie assigned to one of three genres: Action, Comedy, or Drama, and represented through a bag-of-words model derived from plot-related keywords. To provide a fair and semantically informative evaluation, BERT embeddings are used to initialize node features for semantic entities, while additional paper–paper and movie–movie connections are constructed based on metapath proximity and cosine similarity measures. The effectiveness of the proposed framework is evaluated using two standard clustering metrics: Normalized Mutual Information (NMI), which measures the agreement between predicted communities and ground-truth labels, and Adjusted Rand Index (ARI), which assesses clustering similarity while accounting for random assignments.

*4.1. Quantitative Results*

*4.1.1. Structural Graph Analysis*

We analyze the structural properties of the experimental DBLP graph to improve interpretation of quantitative results. We constructed a paper–paper relational network consisting of 68,639,530 edges: a graph density of about 0.332 and a sparsity value of 0.668, which indicate that the structure is not too dense and neighborhood aggregation and information propagation in the entire graph can still be efficient. Moreover, the network has a homophily score of about 0.325, which means that papers belonging to similar research domains are reasonably connected through common authors, venues, or semantic relationships. This shows that, while a moderate level of homophily exists, structural connectivity is not enough to achieve good accuracy in community detection, and thus the need to combine semantic representation learning to learn node features with heterogeneous graph structural modeling.

**Table 2.**
Comparative Performance of Community Detection Methods on the DBLP and IMDB Datasets.

| **Dataset** | **DBLP** | | **IMDB** | |
|---|---|---|---|---|
| **Metric** | **NMI** | **ARI** | **NMI** | **ARI** |
| HAN | 0.7912 | 0.8476 | 0.1087 | 0.1001 |
| MAGNN | 0.8081 | 0.8554 | 0.1558 | 0.1674 |
| HAESF | 0.7612 | 0.8051 | 0.0430 | 0.0548 |
| DMGI | 0.5540 | - | 0.1960 | - |
| BP-GCN | 0.6934 | 0.7626 | 0.2298 | 0.1135 |
| **Our (Hybrid-HGT)** | **0.8216** | **0.8782** | **0.4417** | **0.3044** |

*4.1.2. Comparative Analysis*

The quantitative comparison between the proposed Hybrid-HGT framework and state-of-the-art heterogeneous graph learning methods on DBLP and IMDB using Normalized Mutual Information (NMI) and Adjusted Rand Index (ARI) is summarized in **Table 2**. On DBLP, HAN gets NMI and ARI values of 0.7912 and 0.8476, respectively, while MAGNN marginally improves them to 0.8081 and 0.8554 for NMI and ARI, respectively. The multi-layer dense representation method achieves significantly higher performance than DMGI (NMI: 0.5540) and also performs on par with HAESF and BP-GCN. In comparison, the Hybrid-HGT framework achieves NMI and ARI values of 0.8216 and 0.8782, respectively, demonstrating the capability to model both semantic and structural correlations in academic heterogeneous networks. For the harder IMDB dataset, due to the increase in semantic ambiguity and graph connectivity loss, all baseline models drop in performance significantly. HAN has NMI = 0.1087 and ARI = 0.1001, while MAGNN produces 0.1558 and 0.1674, respectively. Among the algorithms, the highest performance is produced by the baseline BP-GCN, with NMI and ARI of 0.2298 and 0.1135, respectively, while HAESF produces the lowest performance, with NMI 0.0430. In spite of the difficulty of this dataset, Hybrid-HGT obtains a significant improvement, with NMI: 0.4417 and ARI: 0.3044, over all competing approaches by big margins. Our results show that our joint use of heterogeneous transformer attention, contrastive representation learning, iterative pseudo-labeling, and structure-aware edge refinement results in significant community detection performance gains over state-of-the-art methods across both dense and sparse heterogeneous graph scenarios.

*4.1.3. Error Analysis and Limitations*

Although our proposed framework is shown to perform well, a few clustering ambiguities may occur in cases of very similar semantic or structural profiles between communities. Networked systems in which every node belongs with different intensity to a range of topics frequently yield conflated communities; overlapping topical orientations or similar patterns of interaction will cause the model to ascribe interactions from many sources as only one community. This can be well understood due to the fact that in heterogeneous setups, dense interconnections and semantic redundancy make the obscurity of clear cluster boundaries an inherent problem for community detection.

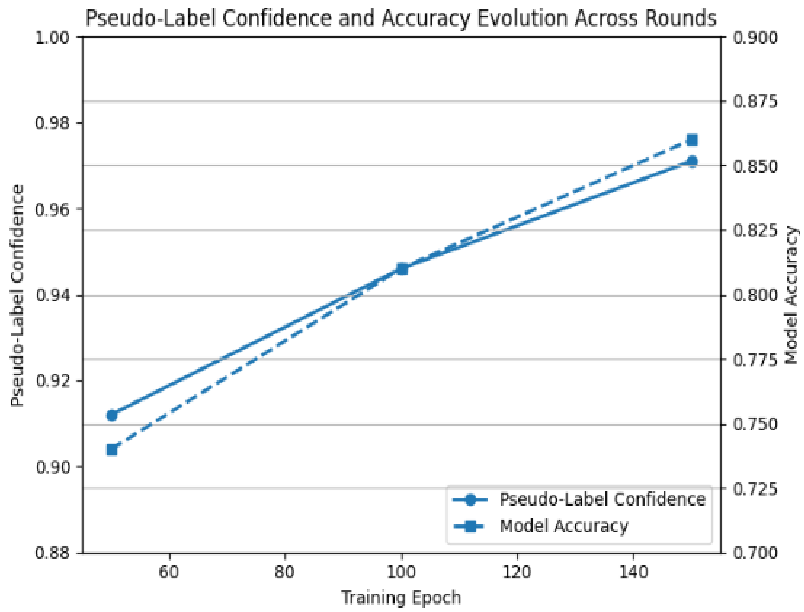


**Figure 1.**
Evolution of Pseudo-Label Confidence and Model Accuracy Across Training Rounds.

The main limitation remains the use of fixed hyperparameters to weight different loss functions, which is an important restriction of the present design. Although the integrated components show efficacy as a whole, there are no sensitivity analyses or exhaustive ablation studies performed to understand the specific individual and combined contributions of each module (e.g., contrastive learning versus clustering-guided optimization). Filling those gaps through targeted experimentation would not only help clarify the extent of influence each component has individually but also yield valuable information regarding the model's robustness as well as stability over different relational structures.

*4.1.4. Pseudo-Label Quality Evolution Analysis*

To evaluate the validity of our multi-round pseudo-labeling strategy, we investigate how pseudo-label confidence and model accuracy evolve at each training round (shown in Figure 1). The outcomes indicate a continuous enhancement in the two measures while training occurs from epoch 50 to epoch 150. More specifically, confidence in pseudo-labels rises from around 0.91 to 0.97, while model prediction accuracy increases from about 0.74 to 0.86. Such a gradual improvement indicates that these pseudo-labels, produced at later training stages, are more accurate and aligned with the actual community structure. The correlation observed between confidence and accuracy also implies that integrated, iterative pseudo-label refinement directly strengthens representation learning, which ultimately allows for higher-quality clustering. Altogether, these results validate the merit of our self-supervised objective, which effectively alleviates label scarcity and progressively improves community detection performance during training.


*DOI: 10.55214/2576-8484.v10i8.13428*

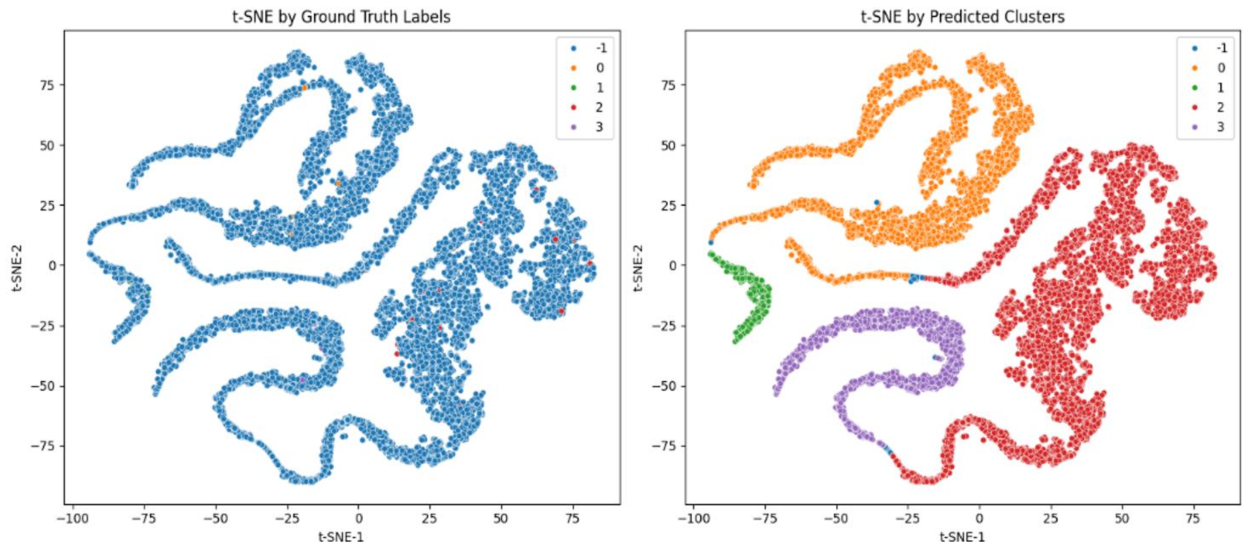


**Figure 2.**
Visualization of the learned node embeddings on the DBLP dataset, comparing the sparse ground-truth labels (left) with the community clusters predicted by the Hybrid-HGT model (right). The predicted assignments reveal four distinct clusters, while −1 denotes unlabeled nodes.

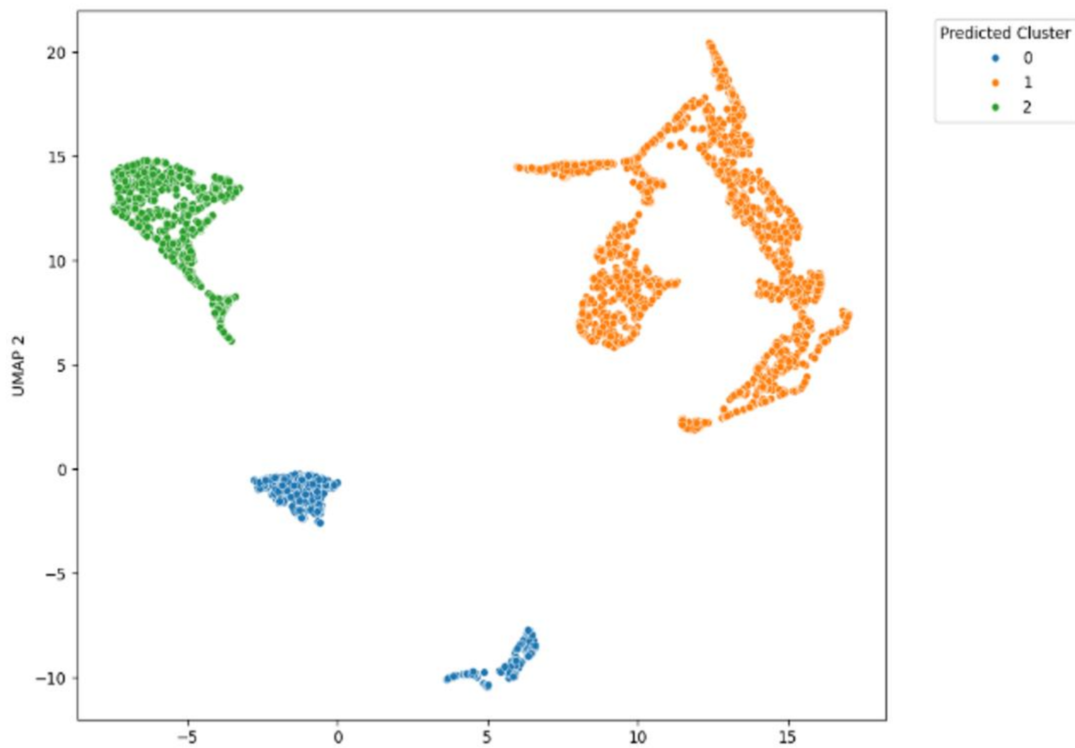


**Figure 3.**
Projection of the latent node embeddings learned by Hybrid-HGT on the IMDB dataset. Colors indicate the three predicted communities, showing clear inter-cluster separation while preserving distinct within-cluster subgroup structures.

### *4.2. Qualitative Evaluation*

We leverage t-SNE projections to qualitatively analyze the latent node embeddings learned by our proposed Hybrid-HGT framework, giving an intuitive way of seeing how well our model captures community structures in heterogeneous networks by mapping high-dimensional embeddings into a 2-D latent space. The embeddings obtained from this step form a gradual transformation of the DBLP label distributions depicted in Figure 2 after Hybrid-HGT training, switching from partially overlapping to well-established, compact community clusters. Some research domains, most notably Artificial Intelligence and Information Retrieval, display partial overlap in ground-truth visualization because they share semantic context (i.e., textual similarities), though this attribution appears weak based on topic-boundary alignment. In contrast, the predicted embeddings generated by the proposed framework show sharper cluster boundaries and more cohesive subgroup structures, even in densely connected graph regions, reflecting that the model has learned to jointly leverage heterogeneous structural information and semantic relationships for improving clustering quality. The t-SNE projections for the IMDB data with true genre labels can be found in Figure 3. Although the framework generates meaningful semantic patterns, some overlap can still be observed from the visualisation, as suggested by lower NMI and ARI scores on the IMDB data in comparison with DBLP. This observation is an inherent limitation of single-label community detection on heterogeneous networks, where semantic categories inherently overlap, simply because, in the real-world movie data, films cannot usually be strictly categorized into a single-community genre due to sharing features that are tied to multiple genres. These results therefore imply some interesting future avenues of research, such as extending the framework toward multi-label community detection in which nodes may belong to multiple communities at once and hierarchical community detection methods that can capture nested semantic structures in heterogeneous networks. In summary, the qualitative visualizations validate the effectiveness of Hybrid-HGT in accurately identifying community structures while also maintaining intra-community tightness and increasing inter-community separability on heterogeneous networks.

## 5. Conclusion and Future Work

This work presents Hybrid-HGT, an end-to-end model for heterogeneous attributed graph community detection that integrates heterogeneous graph transformers, contrastive learning, clustering-guided optimization, and multi-round pseudo-labeling into a single unified architecture. While HGT provides expressive potential to capture complex semantic and structural relationships across the heterogeneous graph, the contrastive learning objective contributes to enhancing both the robustness and discriminative capacity of the learned embeddings. By incorporating clustering refinement and iterative pseudo-label propagation, we further demonstrate that our model is able to achieve community preservation in unsupervised settings where there are limited numbers of labeled data. Experimental results on the DBLP and IMDB benchmark datasets showed that Hybrid-HGT was superior compared to several state-of-the-art heterogeneous graph learning methods, with higher clustering performance in terms of NMI and ARI scores, as well as clearer community separability in qualitative visualizations. Overall, these results also indicate that the framework retains strong semantic cohesion and structural interpretability in challenging environments with sparse connectivity and semantic ambiguity (e.g., IMDB). Despite these improvements, however, some overlap between semantically related communities is still observed, which is not surprising since this problem remains inherently challenging due to the uncertainty involved in separating closely related groups in heterogeneous networks. However, this limitation indicates that extending the framework toward overlapping, hierarchical, or multi-label community detection settings may be particularly interesting, as at least some nodes could naturally belong to multiple communities at once. The adaptation of Hybrid-HGT to dynamic and temporally evolving heterogeneous graphs is an interesting direction for future work, which can help to better model community evolution in real-world systems (e.g., social networks, recommendation platforms, as well as citation networks). Moreover, plugging in large-scale self-supervised pretraining methodologies and higher-order semantic refinement approaches could provide advantages in scalability, transferability, and robustness on out-of-distribution graph structures. All in all, Hybrid-HGT provides a flexible and semantically infused basis for next-generation methods of community detection in complex heterogeneous networks.

## Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

## Copyright:



## References

[1] S. Souravlas, A. Sifaleras, M. Tsintogianni, and S. Katsavounis, "A classification of community detection methods in social networks: A survey," *International Journal of General Systems*, vol. 50, no. 1, pp. 63-91, 2021. https://doi.org/10.1080/03081079.2020.1863394

[2] F. R. Khawaja, Z. Zhang, Y. Memon, and A. Ullah, "Exploring community detection methods and their diverse applications in complex networks: A comprehensive review," *Social Network Analysis and Mining*, vol. 14, no. 1, p. 115, 2024. https://doi.org/10.1007/s13278-024-01274-1

[3] P. Chunaev, "Community detection in node-attributed social networks: A survey," *Computer Science Review*, vol. 37, p. 100286, 2020. https://doi.org/10.1016/j.cosrev.2020.100286

[4] V. Moscato and G. Sperlì, "A survey about community detection over On-line Social and Heterogeneous Information Networks," *Knowledge-Based Systems*, vol. 224, p. 107112, 2021. https://doi.org/10.1016/j.knosys.2021.107112

[5] C. Liu, J. Wu, W. Liu, and W. Hu, "Enhancing graph neural networks by a high-quality aggregation of beneficial information," *Neural Networks*, vol. 142, pp. 20-33, 2021. https://doi.org/10.1016/j.neunet.2021.04.025

[6] M. K. Najafabadi, R.-A. Chen, J. Rezazadeh, A. Beheshti, and N. Shabani, "From theory to practice: The evolution and comparative analysis of homogeneous vs. heterogeneous Graph Neural Networks in recommender systems," *Neurocomputing*, vol. 624, p. 129446, 2025. https://doi.org/10.1016/j.neucom.2025.129446

[7] F. Liu *et al.*, "Deep learning for community detection: Progress, challenges and opportunities," *arXiv preprint arXiv:2005.08225*, 2020. https://doi.org/10.48550/arXiv.2005.08225

[8] X. Zhang, H. Liu, Q. Li, X.-M. Wu, and X. Zhang, "Adaptive graph convolution methods for attributed graph clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 12, pp. 12384-12399, 2023. https://doi.org/10.1109/TKDE.2023.3278721

[9] G. Cui, J. Zhou, C. Yang, and Z. Liu, "Adaptive graph encoder for attributed graph embedding," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020.

[10] E. Dai *et al.*, "A comprehensive survey on trustworthy graph neural networks: Privacy, robustness, fairness, and explainability," *Machine Intelligence Research*, vol. 21, no. 6, pp. 1011-1061, 2024. https://doi.org/10.1007/s11633-024-1510-8

[11] Z. Hu, Y. Dong, K. Wang, and Y. Sun, "Heterogeneous graph transformer," in *Proceedings of the Web Conference 2020*, 2020.

[12] X. Fu, J. Zhang, Z. Meng, and I. King, "Magnn: Metapath aggregated graph neural network for heterogeneous graph embedding," in *Proceedings of the Web Conference 2020*, 2020.

[13] C. Park, D. Kim, J. Han, and H. Yu, "Unsupervised attributed multiplex network embedding," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020.

[14] X. Wang, D. Bo, C. Shi, S. Fan, Y. Ye, and P. S. Yu, "A survey on heterogeneous graph embedding: Methods, techniques, applications and sources," *IEEE Transactions on Big Data*, vol. 9, no. 2, pp. 415-436, 2022. https://doi.org/10.1109/TBDATA.2022.3177455

[15] F. Chen, Y.-C. Wang, B. Wang, and C.-C. J. Kuo, "Graph representation learning: a survey," *APSIPA Transactions on Signal and Information Processing*, vol. 9, p. e15, 2020. https://doi.org/10.1017/ATSIP.2020.13

[16] Y. Zhao, W. Li, F. Liu, J. Wang, and A. M. Luvembe, "Integrating heterogeneous structures and community semantics for unsupervised community detection in heterogeneous networks," *Expert Systems with Applications*, vol. 238, p. 121821, 2024. https://doi.org/10.1016/j.eswa.2023.121821

[17] G. Zhou and R.-F. Wang, "The heterogeneous network community detection model based on self-attention," *Symmetry*, vol. 17, no. 3, p. 432, 2025. https://doi.org/10.3390/sym17030432

[18] X. Wang *et al.*, "Heterogeneous graph attention network," in *The World Wide Web Conference*, 2019.

[19] G. Wu, J. Jiang, and X. Liu, "A practical contrastive learning framework for single-image super-resolution," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 11, pp. 15834-15845, 2023. https://doi.org/10.1109/TNNLS.2023.3290038

[20] B. Fanseu Kamhoua, L. Zhang, K. Ma, J. Cheng, B. Li, and B. Han, "Grace: A general graph convolution framework for attributed graph clustering," *ACM Transactions on Knowledge Discovery from Data*, vol. 17, no. 3, pp. 1-31, 2023. https://doi.org/10.1145/3544977

[21] P. Veličković, W. Fedus, W. L. Hamilton, P. Liò, Y. Bengio, and R. D. Hjelm, "Deep graph infomax. Apollo - University of Cambridge Repository," 2018. https://doi.org/10.17863/CAM.40744

[22] W. Che, Z. Liu, Y. Wang, and J. Liu, "Merit: Multi-level graph embedding refinement framework for large-scale graph," *Complex & Intelligent Systems*, vol. 10, no. 1, pp. 1303-1318, 2024. https://doi.org/10.1007/s40747-023-01211-3

[23] S. H. Lee, F. Ji, and W. P. Tay, "Learning on heterogeneous graphs using high-order relations," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 3175-3179). IEEE*, 2021.

[24] Z. Sun, "Robustness, heterogeneity and structure capturing for graph representation learning and its application," Doctoral Thesis. Durham University, Durham, United Kingdom, 2024.



*ISSN: 2576-8484*
*Vol. 10, No. 8: 36-47, 2026*
*DOI: 10.55214/2576-8484.v10i8.13428*