



# Matching Biomedical Ontologies via a Hybrid Graph Attention Network

Peng Wang<sup>1,2\*</sup> and Yunyan Hu<sup>1</sup>

<sup>1</sup>School of Computer Science and Engineering, Southeast University, Nanjing, China, <sup>2</sup>School of Cyber Science and Engineering, Southeast University, Nanjing, China

Biomedical ontologies have been used extensively to formally define and organize biomedical terminologies, and these ontologies are typically manually created by biomedical experts. With more biomedical ontologies being built independently, matching them to address the problem of heterogeneity and interoperability has become a critical challenge in many biomedical applications. Existing matching methods have mostly focused on capturing features of terminological, structural, and contextual semantics in ontologies. However, these feature engineering-based techniques are not only labor-intensive but also ignore the hidden semantic relations in ontologies. In this study, we propose an alternative biomedical ontology-matching framework BioHAN via a hybrid graph attention network, and that consists of three techniques. First, we propose an effective ontology-enriching method that refines and enriches the ontologies through axioms and external resources. Subsequently, we use hyperbolic graph attention layers to encode hierarchical concepts in a unified hyperbolic space. Finally, we aggregate the features of both the direct and distant neighbors with a graph attention network. Experimental results on real-world biomedical ontologies demonstrate that BioHAN is competitive with the state-of-the-art ontology matching methods.

## OPEN ACCESS

**Edited by:**

Yucong Duan,  
Hainan University, China

**Reviewed by:**

Jiang Bian,  
University of Florida, United States  
Xingsi Xue,  
Fujian University of Technology, China

**\*Correspondence:**

Peng Wang  
pwang@seu.edu.cn

**Specialty section:**

This article was submitted to  
Computational Genomics,  
a section of the journal  
*Frontiers in Genetics*

**Received:** 10 March 2022

**Accepted:** 20 June 2022

**Published:** 22 July 2022

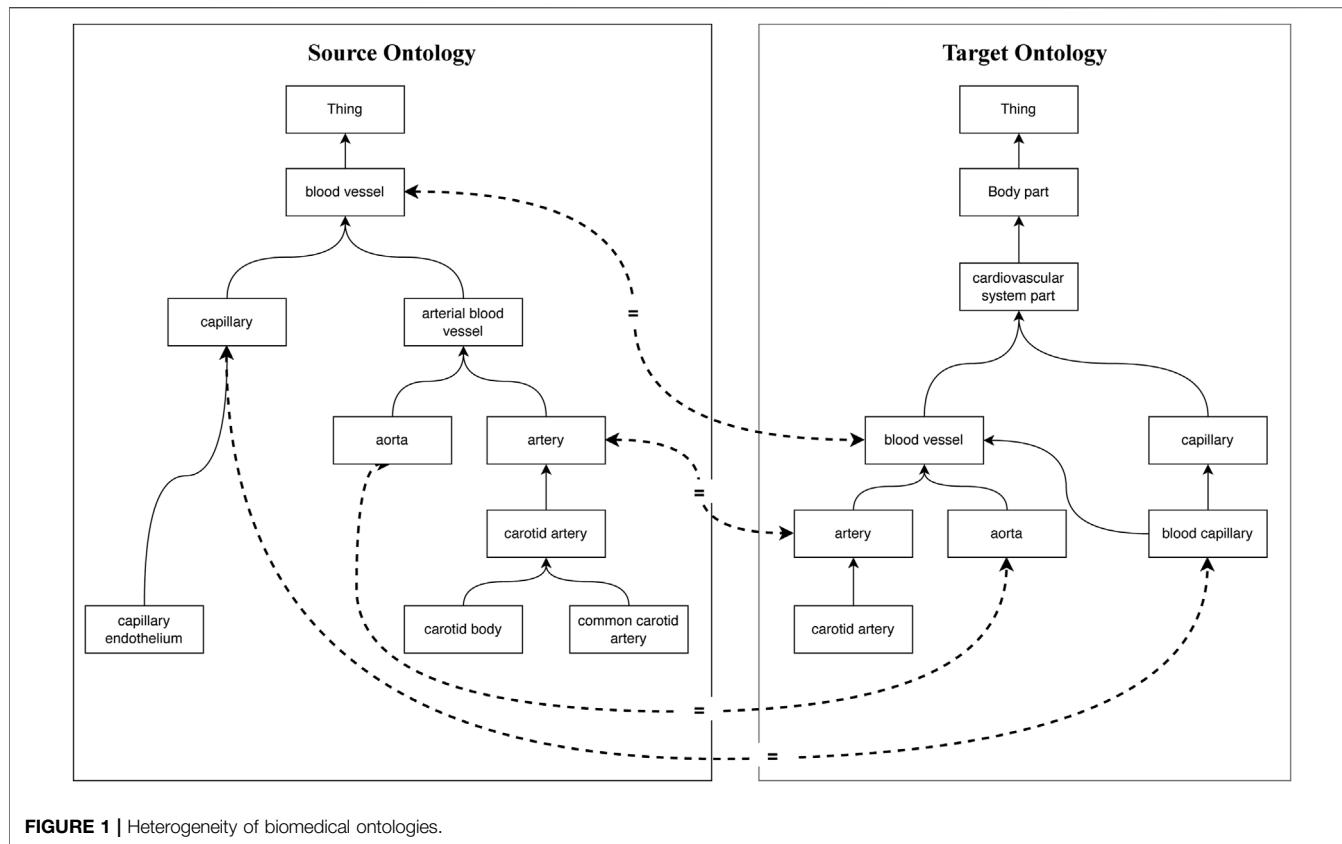
**Citation:**

Wang P and Hu Y (2022) Matching Biomedical Ontologies via a Hybrid Graph Attention Network. *Front. Genet.* 13:893409.  
doi: 10.3389/fgene.2022.893409

**Keywords:** biomedical ontology, ontology matching, graph attention network, embedding, hyperbolic attention

## 1 INTRODUCTION

Ontology is an explicit, interoperable, extensible, scalable, and formal definition to describe knowledge as a set of domain vocabularies that contain concepts, relations between concepts, and individuals of concepts (Ramis et al., 2014). In past decades, various biomedical ontologies, such as the National Cancer Institute Thesaurus (NCI) (Golbeck et al., 2003), Foundation Model of Anatomy (FMA) (Rosse and Mejino, 2003), Systemized Nomenclature of Medicine (SNOMED-Clinical Terms [SNOMED-CT]) (Donnelly et al., 2006), and Uberon (Mungall et al., 2012) have been widely used for medical data format standardization (Cimino and Zhu, 2006), medical or clinical knowledge representation and integration (Isen et al., 2012), and medical decision making (De Potter et al., 2012) to provide standard semantics. With the continuous evolution of biomedical data, biomedical vocabularies have become complicated and ambiguous, which leads to challenges in developing biomedical applications. Moreover, new biomedical ontologies are constructed independently with diverse ways of defining overlapping biomedical terminologies or components, which also leads to more heterogeneity (Xie et al., 2016). As shown in Figure 1, the entities are connected via the *subClassOf* relation, and the equivalent concepts are linked via dotted lines. It can be found that for the same concept, “blood vessel” in the source and target ontologies, they are organized and interpreted at different levels of granularity, named conceptual heterogeneity. In addition, the concepts that share the same morphology “capillary” indicate



**FIGURE 1 |** Heterogeneity of biomedical ontologies.

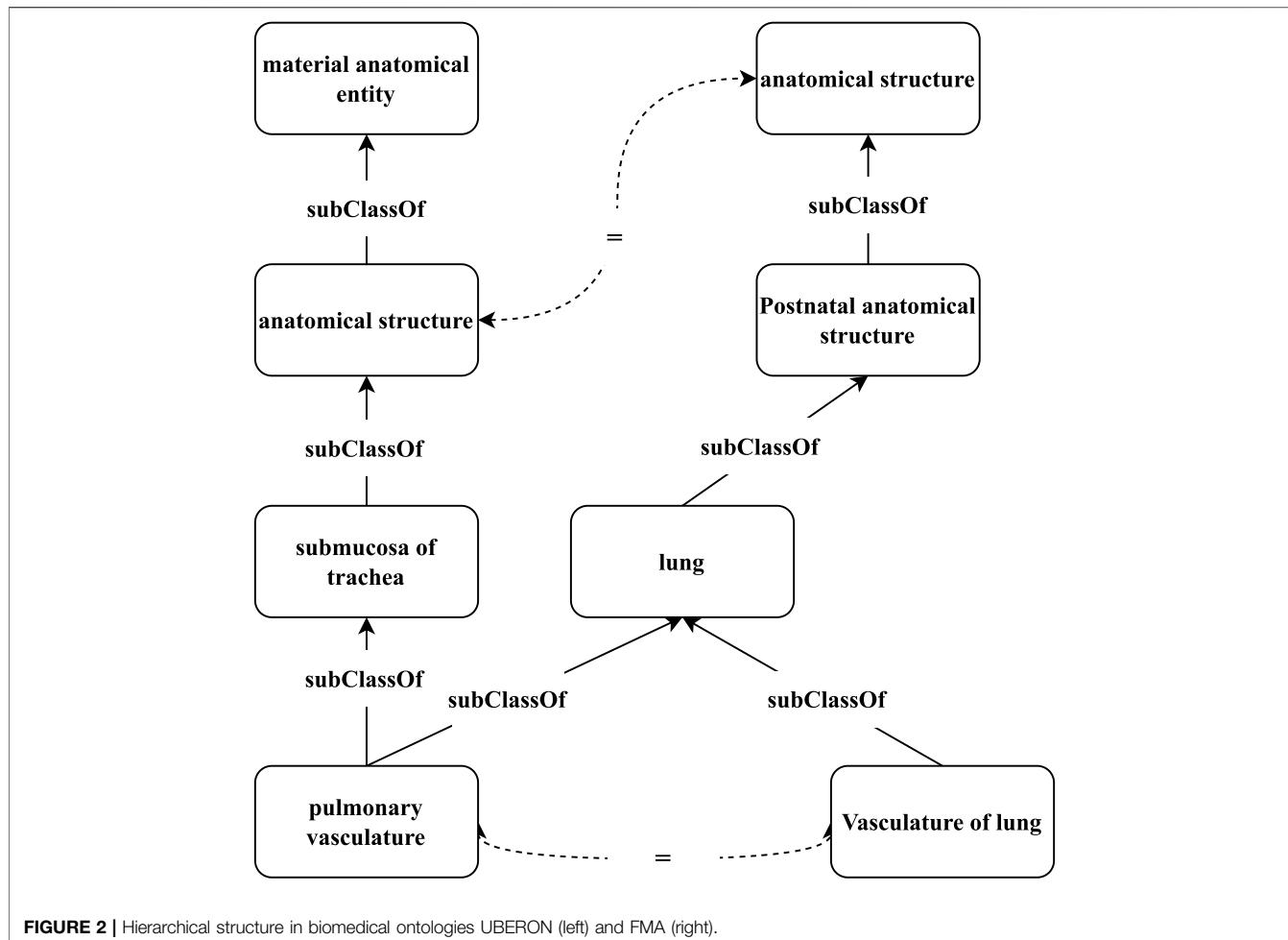
different semantics in different ontologies, which is called semiotic heterogeneity. To implement interoperability across biomedical ontologies, discovering semantic relations between them is critically important (Xue, 2020). Ontology matching is a key technique to find semantic correspondences between the elements of different ontologies to achieve interoperability (Shvaiko and Euzenat, 2011).

Most existing ontology matching methods have focused on extracting features from terminological, structural, extensional (individuals of concepts) information, and external resources (Nezhadi et al., 2011; Otero-Cerdeira et al., 2015; Babalou et al., 2016; Chauhan et al., 2018). They use logical reasoning and rule-based techniques to extract sophisticated features, which are then used to compute the similarities of ontological elements (i.e., concepts, properties, and individuals) that promote ontology matching.

These feature-based methods (e.g., AML (Faria et al., 2013), FCA\_Map (Zhao et al., 2018), LogMap (Jiménez-Ruiz and Cuenca Grau, 2011), and XMap (Djeddi and Khadir, 2014)) elaborate features of data to evaluate element similarity and derive semantic correspondences. However, the features in one ontology usually cannot be transferred to others. Consequently, the effectiveness and generality of those ontology matching methods vary significantly (Kolyvakis et al., 2018).

Recently, graph-based representation learning (Kipf and Welling, 2016; Hamilton et al., 2017) has become a powerful model for learning vector representations of graph-structured

data. In graph neural networks (GNNs), the representation of a node is learned through recursively aggregating the representations of its local neighboring structure and propagation of features from neighboring nodes. Several studies (Chen et al., 2017; Wang Z et al., 2018; Wu et al., 2019; Sun et al., 2020) exploit GNNs for embedding-based matching in knowledge graphs (KGs), and have achieved promising results. However, existing GNN-based matching models still face some problems in ontology matching. First, ontology matching may face semantic imbalance because the distributions and amounts of semantic descriptions in different ontologies are generally different. We argue that if we can enrich the ontologies by using the metadata, given axioms, and auxiliary descriptions from external domain resources, and incorporate a rich set of semantic relationships, the derived ontologies can be matched with higher precision and recall. To overcome this problem, we consider designing an ontology-enriching method. Second, a distinguishable characteristic of biomedical ontologies, compared to open-domain knowledge bases such as YAGO (Suchanek et al., 2007), Wikidata (Vrandečić and Krötzsch, 2014), and DBpedia (Lehmann et al., 2015), is their domain specificity. These biomedical ontologies often have rich hierarchical structures that systematically organize biomedical concepts into categories and subcategories from general to specific. **Figure 2** shows an example of a hierarchical structure in different biomedical ontologies. The hierarchical structures of the corresponding



**FIGURE 2 |** Hierarchical structure in biomedical ontologies UBERON (left) and FMA (right).

pairs in different ontologies are similar to some extent. For example, the hierarchy (through *subClassOf* relation) of “pulmonary vasculature” in UBERON and “Vasculature of lung” in FMA is similar, whereas the terminologies are morphologically different. Therefore, capturing such hierarchical structures would be useful for identifying aligned concepts and improving the matching performance. Finally, since different ontologies usually have heterogeneous schemas and incompleteness (Schneider and Simkus, 2020), the matching pairs usually have some dissimilar neighboring structures. Even though we assume that the ontologies to be matched are complete, because of the schema heterogeneity, the non-isomorphism in the neighboring structures from different ontologies is still inevitable. As shown in **Figure 2**, the one-hop neighbors of the matching pair (“pulmonary vasculature” and “Vasculature of lung”) are different, while they share the same distant neighbor “anatomical structure.” Motivated by the phenomenon that the relevant information could appear in both direct and distant neighbors of matching concepts, the aggregated structural semantics of a concept should include not only its local neighbors, but also the related distant neighbors. In addition, to keep the matching performance, we use an attention mechanism to realize the semantic relatedness of different

neighbors, which could further discover and aggregate important neighbors.

To address these issues, we propose a biomedical ontology matching framework, BioHAN, with a hybrid graph attention network. The underlying idea is to first enrich and refine the ontologies to be matched with the given axioms and auxiliary semantic descriptions from external resources, such as UMLS (Bodenreider, 2004). Then, the neighborhood information is aggregated within multiple hops in the enriched ontologies, capturing both local and global features, into hyperbolic representations that are complementary to each other. Both representations are jointly optimized to improve ontology matching performance. The main contributions of this study are listed as follows:

- We propose a matching method BioHAN for biomedical ontologies. BioHAN first enriches the ontologies for matching via the axioms and logical rules. Then it further learns the representations with the hierarchical structure to realize ontology matching.
- We propose a lightweight and effective way to enrich and refine ontology with the metadata, axioms, and auxiliary semantic information from external resources, which is

- helpful to discover and simplify the hidden and implicit semantics in ontologies.
- To capture the hierarchical features in an ontology, we leverage hyperbolic graph convolution layers to encode the parent and child concepts in the hyperbolic space.
  - To further address the heterogeneity and better capture the semantics of concepts, we introduce an attention mechanism to weigh different neighbors and incorporate multi-hop neighbors to learn both the local and global hierarchical structures.
  - We implement our proposed matching method and conduct systematic experiments on biomedical ontologies datasets. The evaluation of the Ontology Alignment Evaluation Initiative 2021 (OAEI 2021) shows that our method achieves significantly promising results.

The study is structured as follows. In **Section 2**, we describe relevant preliminaries of ontology matching and the overview of our proposed method. In **Section 3**, we illustrate the ontology-enriching operation, including ontology preprocessing and augmenting. In **Section 4**, the implementation details of our proposed matching method BioHAN are presented. **Section 5** describes our experiments, the results, and the experimental analysis and discussion. In **Section 6**, related work about ontology matching is systematically reviewed and introduced. **Section 7** summarizes our main findings, and presents perspectives on future work.

## 2 PRELIMINARIES AND METHOD OVERVIEW

### 2.1 Ontology Matching

Let  $\mathcal{C}$  be the set of concepts,  $\mathcal{R}$  be the set of relations, and  $\mathcal{T} = \mathcal{C} \times \mathcal{R} \times \mathcal{C}$  be the set of triples or statements, then a biomedical ontology can be represented as  $O = (\mathcal{C}, \mathcal{R}, \mathcal{T})$ . The matching between two ontologies  $O_s$  and  $O_t$  is  $\mathcal{M} = \{m_k | m_k = \langle e_i, e_j, r, s \rangle\}$  (Euzenat and Shvaiko, 2007), where  $\mathcal{M}$  is an alignment;  $m_k$  is a correspondence  $\langle e_i, e_j, r, s \rangle$ ;  $e_i$  and  $e_j$  are elements from  $O_s$  and  $O_t$ , respectively;  $r$  is the semantic relation between  $e_i$  and  $e_j$ ; and  $s \in [0, 1]$  is the confidence about a correspondence. Therefore, an alignment  $\mathcal{M}$  is a set of correspondences  $m_k$ .

### 2.2 Graph Neural Networks

Graph neural networks (GNNs) are effective for various applications with graph-structured data (Zhou et al., 2020). A GNN framework usually has a graph encoder and a graph decoder, and its input is an adjacency matrix and features nodes and edges. The encoder uses the graph structure to propagate and aggregate information across nodes, and learns embeddings for local structure. A graph decoder is often used to compute similarity scores for all node pairs. Depending on the graph properties and aggregation strategies, some GNN frameworks have been proposed.

The vanilla GCN is a popular variant of the GNN (Kipf and Welling, 2016), in which the hidden representation of node  $i$  at the  $l$ -th ( $l > 0$ ) layer  $h_i^{(l)}$  is computed as

$$\mathbf{h}_i^{(l)} = \sigma \left( \sum_{j \in \mathcal{N}_i \cup \{i\}} \frac{1}{c_{ij}} \mathbf{W}_i^{(l)} \mathbf{h}_i^{(l-1)} \right) \quad (1)$$

where  $\sigma(\cdot)$  is an activation function;  $\mathbf{W}^{(l)}$  is the weight matrix of the  $l$ -th layer and  $c_{ij}$  is for normalization; and  $\mathcal{N}_i$  denotes the neighbor set of node  $i$ . The vanilla GCN encodes node  $i$  as the mean pooling of the representations of its neighbors and node  $i$  itself from the last layer. The input vector fed to the first layer is denoted as  $h_i^{(0)}$ .

A graph attention network (GAT) (Veličković et al., 2018) is a novel convolution-style neural network with masked self-attention layers. In contrast to the GCN, it allows for implicitly setting different weights to nodes of the same neighboring node. Moreover, analyzing the learned attention weights could improve interpretability. Formally, the attention weight  $\alpha_{ij}^{(l)} \in \mathcal{R}$  between  $i$  and  $j$  at the  $l$ -th layer is computed as follows:

$$\alpha_{ij}^{(l)} = \frac{\exp(\text{LeakyReLU}(\mathbf{a}^T [\mathbf{W}\mathbf{h}_i \| \mathbf{W}\mathbf{h}_j]))}{\sum_{j \in \mathcal{N}_i} \exp(\text{LeakyReLU}(\mathbf{a}^T [\mathbf{W}\mathbf{h}_i \| \mathbf{W}\mathbf{h}_j]))} \quad (2)$$

Here,  $\cdot^T$  denotes transposition;  $\mathbf{a}$  is an attention weight matrix;  $\|$  is the concatenation operation; and  $\text{LeakyReLU}$  is used to achieve nonlinear transformation.

### 2.3 Method Overview

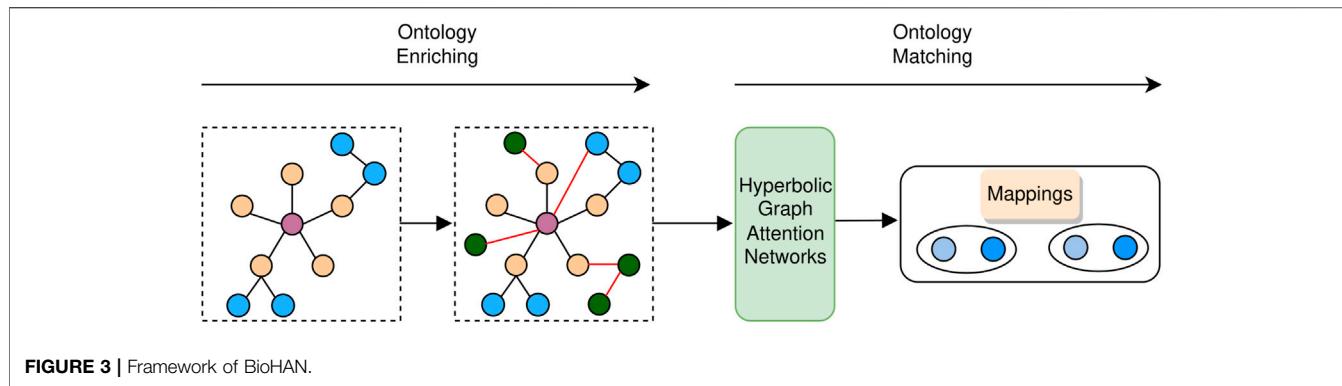
As shown in **Figure 3**, our proposed BioHAN comprises two phases: ontology enriching and ontology matching. Given a biomedical ontology, the ontology enriching phase first preprocesses the ontology with the metadata and axioms, which complements the informative representations hidden in the ontology. It also explores matching seeds between the processed ontologies by supplementing some missing semantics through external resources. The ontology matching phase takes as input the derived ontology. Structures of ontologies are captured via graph attention networks for structural representation learning. Moreover, the lexical semantics of the concepts in ontologies are used, providing complementary clues for ontology matching.

## 3 ONTOLOGY ENRICHING

In this section, we will discuss ontology preprocessing and augmenting operation to enrich the initial ontology. Specifically, we first preprocess the ontology to discover the hidden semantics and represent them clearly. Then, we use ontology augmentation strategies to enrich the ontologies.

### 3.1 Ontology Preprocessing

We notice that there are two common facts in biomedical ontologies. On the one hand, some semantic information is hidden or unclear, which is expressed by complex axioms or ontology semantics. However, to further understand an ontology, such information is useful. On the other hand, some triples are used to describe the building and version



**TABLE 1 |** Example of ontology collection.

```

< rdf:Description rdf:about = "http://bioontology.org/ontologies/fma/physical
state"
< rdfs:range >
< rdf:List >
< rdf:li > Gas </rdf:li >
< rdf:li > Liquid </rdf:li >
< rdf:li > Semi-solid </rdf:li >
< rdf:li > Solid </rdf:li >
< rdf:List >
</rdfs:range >
</rdf:Description >

```

information for an ontology. These statements simply increase the size of the ontology and are useless for the definitions of concepts and properties. Therefore, we conduct a preprocessing operation to refine ontologies. Specifically, we make the complex expressions of ontologies much simpler and clearer.

For the ontology language RDFS and OWL, they provide mechanisms for describing groups of related resources and the relationships between these resources, where OWL is an extension of RDFS, providing description logic-based primitives with richer expressive ability and stronger reasoning ability. In an ontology, containers (e.g., `rdf:Bag`, `rdf:Seq`, and `rdf:Alt`) and collections (e.g., `rdf:List`) are used to describe a set of resources in RDFS and OWL. They simplify the ontology expressions but hide some indirect semantics. We clearly define the semantics of the members in containers and collections, and then delete those redundant and complex statements. **Table 1** shows the range of property “physical state” through a collection `rdf:List` in RDFS format. Through the RDFS description, we can know that for the property “physical state” in the ontology “<http://bioontology.org/ontologies/fma>,” its values could be one of “Gas,” “Liquid,” “Semi-solid,” and “Solid.” Each value is represented via `rdf:li`. However, the members would be represented as anonymous nodes while parsing the ontology, such as `<physicalstate, range, BN>`, `<BN, range, Liquid>`, where `BN` denotes an anonymous node with no specific meaning. These statements are difficult to understand directly.

Therefore, it is necessary to formulate this implicit knowledge, such as `<physicalstate, range, Liquid>`.

In addition, to further mine the semantic descriptions in the biomedical ontologies, a rule-based reasoning method is proposed to discover the hidden information.

- 1) Enriching domain and range: given a property  $p_a$ , let  $p_b$  be the sub-property of  $p_a$ . Then we can infer that all semantics of the domain and range of  $p_a$  could be inherited by  $p_b$ . According to this rule, the semantics of sub-properties will be defined more comprehensively.
- 2) Enriching the concept axioms: given a concept axiom (e.g. `owl:oneOf`, `owl:intersectionOf`, `owl:unionOf`, `owl:equivalentClass`, etc.), its equivalent semantics could be rewritten by following rules. If a complex concept  $A \sqcap B$  is defined by the axiom `owl:intersectionOf`, where the complex concept has a sub concept  $C$ ,  $A \sqsupset C$  and  $B \sqsupset C$  could be added to the ontology. If one complex concept  $A \sqcup B$  is defined by the axiom `owl:unionOf`, where the complex concept has a super concept  $C$ , so  $C \sqsubset A$  and  $C \sqsubset B$  could be added to the ontology. Similarly, we can also rewrite semantics of `owl:oneOf` and `owl:equivalentClass`. Therefore, complex semantics of concept axioms could be clearly defined.
- 3) Enriching the property axioms: given a property axiom (e.g. `owl:SymmetricProperty`, `owl:TransitiveProperty`, `owl:equivalentProperty`, etc.), relevant semantic extension could be realized by following rules. If a property  $p$  is declared by axiom `owl:SymmetricProperty` and there is a statement `<A, p, B>`, a new statement `<B, p, A>` could be added to the ontology. If a property  $p$  is declared by axiom `owl:TransitiveProperty` and there are statements `<A, p, B>` and `<B, p, C>`, then a new statement `<A, p, C>` could be added to the ontology.
- 4) Enriching `owl:sameAs` axiom: given a statement `<A, owl:sameAs, B>`, then the equivalent individuals  $A$  and  $B$  could share their semantic information.
- 5) Enriching properties in the concept hierarchy: given `<p, rdfs:domain, A>` and `<B, rdfs:subClassOf, A>`, we can infer an implicit statement `<p, rdfs:domain, B>`. According to this rule, the property’s constraints about one concept could be extended to its sub-concepts.

## 3.2 Ontology Augmentation

Even though the derived ontologies have clearly specified the hidden semantics, they are still insufficient to some extent. Some semantic relationships are still missing, which may lead to the sparse problem of ontology structure. To alleviate this problem, we introduce several augmentation heuristics to enrich biomedical ontologies through the external domain resources, that is, UMLS.

### 3.2.1 Concept Augmentation

We first explore the anchors between the ontologies to be matched and the external resources, which is performed by using a simple string-based technique. Then, for one concept in ontologies, the relative semantics (e.g. rdfs:label, owl:annotation, owl:equivalentClass, etc.) of its anchored concept in external resources could be transferred and added to the ontology. Concept augmentation can significantly enrich ontologies with available information from external resources.

### 3.2.2 Neighborhood Augmentation

Relations between source and target concepts could also be derived from the anchored concepts in external resources. Specifically, if there is a relation between concepts  $i$  and  $j$  of the external resource, their anchors  $i'$  and  $j'$  are also linked by this relation. The goal is to reduce the semantic gap between ontologies by adding the missing structural information and solving the problem of sparse ontology graphs.

With the augmented ontologies, our matching framework enables sufficient learning of ontology representations. To match the concepts in ontology  $O_s$  and ontology  $O_t$ , we use graph pooling to obtain the embeddings of concepts. After investigating different graph pooling methods (Hamilton et al., 2017; Ying et al., 2018), we choose mean-pooling to capture information across concept neighbors. Finally, the graph neural networks take the enriched ontologies  $O_s$  and  $O_t$  as input to find the alignments.

## 4 MATCHING METHOD

In this section, we first embed the elements in ontologies to low dimension vectors, and then discuss the hyperbolic graph attention mechanism. Subsequently, we elaborate on the matching computation and the model training in detail.

### 4.1 Embedding

The terminological descriptions of concepts within a biomedical ontology are generally represented by a sequence of words. We leverage deep learning-based embedding methods (Peters et al., 2018; Devlin et al., 2019) to derive a fixed-size terminological description embedding for each concept. In this study, we choose BioBERT, a high-quality medical language model pre-trained on PubMed abstracts and clinical notes (Lee et al., 2020), to encode concepts. Considering the domain specificity of biomedical ontology, the embedding models toward a specific task can provide significant benefits (Alsentzer et al., 2019; Peng et al., 2019), and are much more appropriate than the general pre-training language model. The embeddings are used as the

initial states  $h^{0,E}$  of concepts, where  $E$  indicates the low-dimensional vectors in the Euclidean space.

## 4.2 Hyperbolic Graph Attention

Conventional GNNs typically capture the graph via message propagation to embed nodes into the Euclidean space. However, it could lead to the distortion of hierarchical structures (Nickel and Kiela, 2017). Hence, we transfer the node representations to a hyperbolic embedding space that can better capture the hierarchical characteristics of tree-like ontologies. In this study, we use a specific model, hyperbolic graph attention network (HGAT) (Zhang et al., 2021), which jointly implements both the expressiveness of a GAT and the superiority of hyperbolic geometry in capturing the hierarchical features. Moreover, multi-hop neighbors are also encoded into concepts, to comprehensively consider a broader context of concepts and alleviate the heterogeneity problem. The network architecture is shown in **Figure 4**.

### 4.2.1 Hyperbolic Feature Projection

The hyperbolic graph attention layer first establishes transformation between the tangent (Euclidean) and Poincaré ball, which is carried out by exponential and logarithmic maps. Specifically, we project the vector in a tangent space to a hyperbolic manifold through the exponential map, whereas the logarithmic map reverses the hyperbolic representation back to the Euclidean space. The initial hyperbolic embedding  $h_i^{0,H}$  of node  $i$  is

$$h_i^{0,H} = \exp_o^K(0, h_i^{0,E}) \quad (3)$$

where  $K$  determines the constant negative curvature  $-1/K(K > 0)$  and the tangent space is centered at point  $o$ . To transform the hyperbolic features from one layer to the next layer, we follow the following computation:

$$h_i^{l,H} = (\mathbf{W}^l \otimes^{K_{l-1}} h_i^{l-1,H}) \oplus^{K_{l-1}} \mathbf{b}^l \quad (4)$$

where  $\otimes$  and  $\oplus$  are hyperboloid matrix multiplication and addition, respectively.

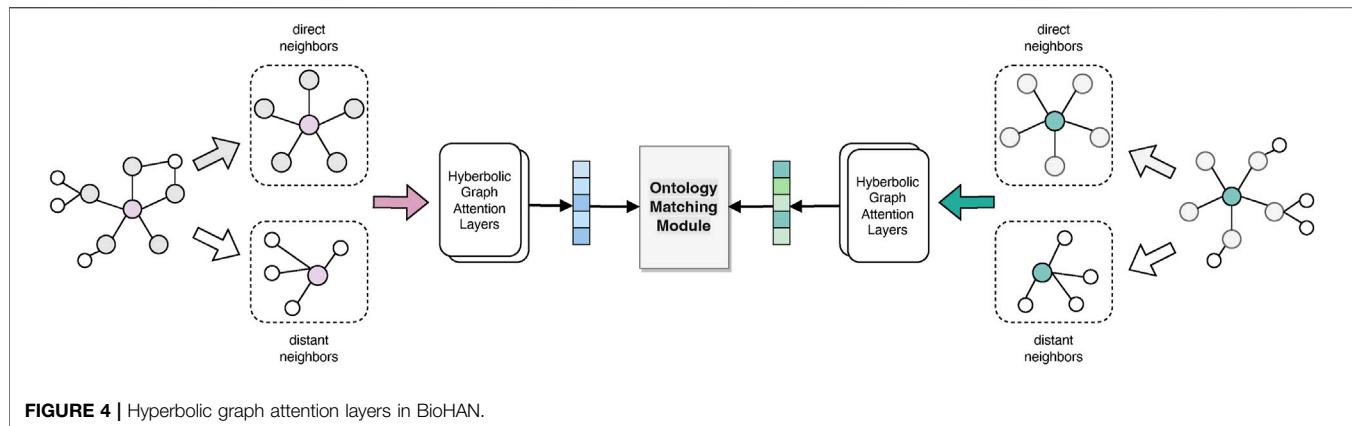
### 4.2.2 Hyperbolic Attention Mechanism

To measure the importance of various neighbors and aggregate the neighbors' features to the center node according to their semantic weights, a self-attention mechanism is performed on the nodes. To that end, one parameterized weight matrix  $\mathbf{W}$  is applied to all nodes to conduct the shared linear transformation. Then, the attention coefficient can be represented with a self-attention weight  $\mathbf{a}$  on the nodes as follows:

$$e_{ij} = \mathbf{a}^T(\mathbf{W}(h_i^h, h_j^h)) \quad (5)$$

$e_{ij}$  indicates the importance of node  $j$  to node  $i$ .

In addition, GAT considers only the local neighbors (i.e., one-hop neighbor nodes) for graph attention, while distant neighboring nodes can also contribute semantics to the central node. To reduce the effects of non-isomorphism in neighboring structures, we introduce distant neighboring information. Without loss of generality, we aggregate both the one-hop and two-hop neighboring information in ontologies, obtaining a proximity matrix.



**FIGURE 4 |** Hyperbolic graph attention layers in BioHAN.

$$P = (B_1 + B_2)/2 \quad (6)$$

where  $B$  is the transition matrix and  $B_k$  denotes the adjacency matrix of  $k$ -th hop.  $B_{ij} = 1/d_i$  if there exists an edge between  $i$  and  $j$  in the  $k$ -th hop, otherwise  $B_{ij} = 0$ . Then,  $P_{ij}$  denotes the topological weight that node  $j$  exerts on  $i$ .

To make coefficients comparable across different concepts, the attention weights are normalized via the softmax function.

$$\alpha_{ij} = \text{softmax}(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}_i} \exp(e_{kj})} \quad (7)$$

Finally, using the topological weights  $P$  and applying the LeakyReLU nonlinearity, the coefficients can be expressed as

$$\alpha_{ij}^{(l)} = \frac{\exp(\text{LeakyReLU}(P_{ij} \cdot \mathbf{a}^T [\mathbf{W}\mathbf{h}_i \parallel \mathbf{W}\mathbf{h}_j]))}{\sum_{k \in \mathcal{N}_i} \exp(\text{LeakyReLU}(P_{ij} \cdot \mathbf{a}^T [\mathbf{W}\mathbf{h}_i \parallel \mathbf{W}\mathbf{h}_j]))} \quad (8)$$

#### 4.2.3 Hyperbolic Attention-Based Aggregation

Similar to GAT, the hyperbolic graph convolution layer aggregates features from a node's local neighbors. There is no notion of a vector space structure in a hyperbolic space, while the hyperboloidal aggregation requires multiplication by a weight matrix along with a bias operation. The main idea is to leverage the logarithmic projection to perform the Euclidean transformation and aggregation in the tangent space, and then transfer the obtained vectors back to the hyperbolic space. In addition, an attention mechanism is applied to learn the semantic relatedness between the neighboring nodes and the central node. Then, the neighbors' features are assembled in accordance with the learned attention coefficients. The hyperbolic attention-based aggregation is defined as follows:

$$\text{AGG}^K(\mathbf{h}^H)_i = \exp\left(\sum_{j \in \mathcal{N}_i} \alpha_{ij} \log_{\mathbf{h}_i^H}(\mathbf{h}_j^{l-1,H})\right) \quad (9)$$

To avoid semantic loss during the information propagation and maintain its transitivity between different convolutional layers, it is also necessary to incorporate the semantics of the central node itself.

$$\text{AGG}^K(\mathbf{h}^H)_i = \exp\left(W_{\text{AGG}}\left(\mathbf{h}_i^H + \sum_{j \in \mathcal{N}_i} \alpha_{ij} \log_{\mathbf{h}_i^H}(\mathbf{h}_j^{l-1,H})\right)\right) \quad (10)$$

where  $W_{\text{AGG}}$  is the aggregated weight matrix, and  $\mathbf{h}_i^H$  is the representation of the central node.

Finally, a non-linear activation function is used to increase the nonlinear expression ability and further improve the performance of the model. Specifically, BioHAN first applies Euclidean non-linear activation in the tangent space and then projects back to the hyperbolic space.

$$\sigma^{\oplus K_{l-1} K_l}(\mathbf{h}^H) = \exp_o^{K_l}(\sigma(\log_o^{K_{l-1}}(\mathbf{h}^H))) \quad (11)$$

The  $l$ -th layers of a hyperbolic graph attention layer are

$$\mathbf{h}_i^{l,H} = \sigma^{\oplus K_{l-1} K_l}(\text{AGG}^{K_{l-1}}(\mathbf{h}^{l-1,H})) \quad (12)$$

where  $-1/K_{l-1}$  and  $-1/K_l$  are the hyperbolic curvatures at the  $(l-1)$ -th and  $l$ -th layer, respectively. After iterative propagation and update of representations between layers, the final hyperbolic vector representations  $\mathbf{h}^H$  can be obtained to represent the concepts.

#### 4.3 Matching

Based on the learned concept representations  $\mathbf{h}^H$  from the hyperbolic graph attention layers, our matching module takes as input pairs of concept embeddings from  $O_s$  and  $O_t$ , and then measures the semantic relatedness with a similarity metric function, defined as follows:

$$\text{sim}(c_i, c_j) = \left\{ \exp\left[\frac{1}{t} \left( d^K(\mathbf{h}_i^H, \mathbf{h}_j^H)^2 - r \right) \right] + 1 \right\}^{-1} \quad (13)$$

where  $d^K(\cdot, \cdot)$  is the hyperbolic distance, and  $r$  and  $t$  are hyperparameters. Then we iteratively match the concepts of two different ontologies using the Stable Marriage algorithm (SM) (Gale and Shapley, 1962) over the concepts' pairwise similarities.

#### 4.4 Training

To improve the matching performance of the proposed method, we jointly consider the reconstruction performance of the

hyperbolic graph attention network and the matching performance of the matching module.

For the hyperbolic graph attention network module, the graph transition matrix of the final output should be as close as possible to the original graph structure. Therefore, the graph reconstruction loss should be minimized.

$$\mathcal{L}^H = \sum_{(i,j) \in E^+} p(c_i, c_j) + \sum_{(i,j) \in E^-} \omega [\mu - p(c_i, c_j)] \quad (14)$$

where  $E^+$  is the set of adjacency concept pairs;  $E^-$  represents the corresponding negative samples;  $\mu$  is the margin value;  $\omega$  is a trade-off factor; and  $[\cdot]_+ = \max(0, \cdot)$ .

Then, for the matching module, we minimize the contrastive matching loss to actualize that the distances between pre-aligned concepts (positive) are as small as possible whereas the unmatched (negative) pairs have a relatively larger distance.

$$\mathcal{L}^M = - \sum_{(i,j) \in \mathcal{M}^+} \log sim(c_i, c_j) - \sum_{(i,j) \in \mathcal{M}^-} \log(1 - sim(c_i, c_j)) \quad (15)$$

where  $\mathcal{M}^+$  is the set of seed correspondences between  $O_s$  and  $O_o$ , and  $\mathcal{M}^-$  denotes the corresponding opposite ones.

The final joint loss function is defined as follows:

$$\mathcal{L} = \mathcal{L}^H + \alpha \cdot \mathcal{L}^M \quad (16)$$

where  $\alpha$  is positive hyper-parameters to control the trade-off among these loss components. The model is trained by minimizing the overall loss and optimized with an Adam (Kingma and Ba, 2014) optimizer.

## 5 EXPERIMENTS

This section reports the experimental results. To verify the effectiveness of BioHAN, we used Python to implement our approaches in Pytorch and conduct the experiments on a computer with an Intel Xeon 4110 CPU, Nvidia 2080Ti GPU, and 64-GB memory.

### 5.1 Datasets

The experiments are performed on the biomedical evaluation benchmark from the Ontology Alignment Evaluation Initiative 2021 (OAEI 2021), which organizes annual evaluation campaigns aiming at evaluating ontology matching technologies. Biomedical ontologies are collected from the Large Biomedical track in OAEI 2021, including the Foundational Model of Anatomy Ontology (FMA), SNOMED CT, and the National Cancer Institute Thesaurus (NCI).

The FMA is an ontology for biomedical informatics that symbolically represents the phenotypic structure of the human body (Rosse and Mejino, 2003). FMA has 78,988 concepts together with 78,985 *isA* triples.

The NCI provides reference terminologies for clinical care, translational and basis research, public information, and administrative activities (Golbeck et al., 2003). It comprises 66,724 concepts and 59,794 *isA* triples.

SNOMED CT is a systematically organized collection of medical terms and provides comprehensive, multilingual clinical healthcare terminology for clinical documentation and reporting (Donnelly et al., 2006). It contains 1,22,222 concepts and 1,05,624 *isA* triples.

The matching tasks are FMA-NCI, FMA-SNOMED, and NCI-SNOMED. On account of the primary hierarchical architecture of ontologies and the deficiency of some other relations, except the hierarchical structure, we mainly consider the *SubClassOf* relationship of these datasets. In this study, we only focus on identifying one-to-one equivalence correspondences between concepts. Seed alignments are extracted from the UMLS (Bodenreider, 2004) and trained as positive samples. The negative alignments are sampled by randomly modifying one of the concepts in the positive sample pairs.

### 5.2 Evaluation Measures

We follow the standard evaluation criteria in OAEI 2021, calculating the precision (P), recall (R), and F1-measure (F1) for each matching task. Given a reference alignment set *Ref* and mapping correspondences *Map*, the precision and recall are calculated as follows:

$$P = \frac{|Map \cap Ref|}{|Map|} \quad (17)$$

$$R = \frac{|Map \cap Ref|}{|Ref|} \quad (18)$$

The F1-measure is the weighted harmonic average of precision and recall, defined as

$$F1 = \frac{2 \times P \times R}{P + R} \quad (19)$$

### 5.3 Experimental Settings

For our proposed BioHAN, each training takes 1,000 epochs with the learning rates among {0.01, 0.001, 0.0001}. The embedding dimension  $d$  is set to 128, and the initial input embedding has the size (d) 512. By default, we stack two hyperbolic graph attention layers in our model. For the hyperbolic graph attention decoder, we set  $r = 2.0$ ,  $t = 1.0$ , and apply trainable curvature, which refer to the parameter configuration in MEDTO (Hao et al., 2021). We set the trade-off hyper-parameters  $\alpha_1$  to 1.0. In addition, for each seed correspondence, we corrupt it and randomly replace it with five additional concepts to generate negative mapping pairs.

### 5.4 Experimental Results

#### 5.4.1 Ontology Matching Results

Table 2 shows the matching results of our proposed model compared with several matching methods or systems based on feature engineering and representation learning. The feature engineering-based top-performing matching systems are

**TABLE 2 |** Results of ontology matching.

| Method          | FMA-NCI      |              |              | FMA-SNOMED   |              |              | SNOMED-NCI   |              |              |
|-----------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
|                 | P            | R            | F1           | P            | R            | F1           | P            | R            | F1           |
| AML             | <b>0.958</b> | 0.910        | <b>0.933</b> | 0.923        | 0.762        | <b>0.835</b> | 0.906        | 0.746        | 0.818        |
| LogMap          | 0.940        | 0.898        | 0.919        | <b>0.941</b> | 0.689        | 0.796        | <b>0.954</b> | 0.667        | 0.785        |
| LogMapBio       | 0.904        | 0.920        | 0.912        | 0.911        | 0.711        | 0.799        | 0.909        | 0.696        | 0.88         |
| MTransE         | 0.627        | 0.640        | 0.633        | 0.505        | 0.475        | 0.490        | 0.254        | 0.378        | 0.304        |
| GCN-align       | 0.813        | 0.783        | 0.798        | 0.763        | 0.729        | 0.746        | 0.745        | 0.775        | 0.760        |
| DAEOM           | 0.882        | 0.689        | 0.774        | 0.719        | 0.693        | 0.706        | 0.891        | 0.682        | 0.773        |
| MEDTO           | 0.944        | 0.874        | 0.908        | 0.871        | 0.762        | 0.813        | 0.901        | <b>0.802</b> | 0.849        |
| BioHAN          | 0.930        | <b>0.922</b> | 0.926        | 0.898        | <b>0.775</b> | 0.832        | 0.911        | 0.797        | <b>0.850</b> |
| BioHAN (w/o OB) | 0.930        | 0.922        | 0.926        | 0.782        | 0.731        | 0.756        | 0.788        | 0.709        | 0.746        |
| BioHAN (w/o HB) | 0.831        | 0.822        | 0.826        | 0.771        | 0.729        | 0.749        | 0.850        | 0.711        | 0.774        |
| BioHAN (w/o AM) | 0.860        | 0.842        | 0.851        | 0.819        | 0.726        | 0.770        | 0.864        | 0.719        | 0.785        |
| BioHAN (w/o MN) | 0.893        | 0.849        | 0.870        | 0.822        | 0.745        | 0.782        | 0.877        | 0.701        | 0.779        |

*Bold* values represents the best results for the column in which they are located.

selected according to the results published in the Large Biomedical track by OAEI 2021. The comparative representation learning models are several recent typical embedding-based entity alignment models (MTransE, GCN-Align) and ontology matching models (DAEOM, MEDTO).

Compared with the extensively developed feature-based approaches such as AML, LogMap, and LogMapBio, our method achieves competitive results across all three tasks. The proposed BioHAN outperforms these rule-based approaches in measure R in FMA-NCI and FMA-SNOMED. AML, LogMap, and LogMapBio heavily rely on lexical features extracted from ontologies, while using representation learning could better capture some hidden semantics to discover more complex matching pairs. We can also observe that entity alignment models (MTransE, GCN-Align) designed for general knowledge bases are insufficient for domain-specific ontology matching. Compared to the representative matching methods (DAEOM, MEDTO), BioHAN also achieves competitive performance. The performance difference between MEDTO and BioHAN validates the importance of hierarchical features. BioHAN explicitly distinguishes and models the hierarchical structure, taking into account both the local and global hierarchical features, and obviously leads to promising results in biomedical ontology matching.

#### 5.4.2 Effectiveness of Ontology Enriching

To evaluate the effectiveness of the enriching phase, we further compare the *isA* triple size during ontology enriching. The detailed statistics concerning the size of each ontology matching task are shown in **Table 3**. Here, *Nodes* means the number of ontology entities, and *isA* is the edges between nodes with the relation *owl: subClassOf* in the ontology graph, while the *origin* and *enriching* represents the change in *isA* triple size before and after the enriching operation.

We can observe that the change in the triple size of both the ontology NCI and SNOMED is explicit, while the FMA remains. The structure of NCI and SNOMED is sophisticated, and contains

**TABLE 3 |** Summary statistics of ontology enriching.

| Ontology | Nodes    | isA (origin) | isA (enriching) |
|----------|----------|--------------|-----------------|
| FMA      | 78,988   | 78,985       | 78,985          |
| NCI      | 66,724   | 59,794       | 75,454          |
| SNOMED   | 1,22,222 | 1,05,624     | 2,03,942        |

substantive *owl:intersectionOf* and *owl:unionOf* property links, especially SNOMED. Specifically, the *owl:intersectionOf* statement describes classes which contain precisely those individuals that are members of the class extension of all class descriptions in the list, while the *owl:unionOf* statement describes an anonymous class containing those individuals occurring in at least one of the class extensions in the list.

Moreover, we compare the matching performance between the proposed BioHAN and its variation BioHAN (w/o OB), which does not pay attention to ontology preprocessing and enriching. Results are also shown in **Table 2**. It is obvious that our model BioHAN consistently outperforms across these tasks, with an average increase of 6.0% in the F1 measure. This is attributed to the critically abundant structural features and implicit semantics added to ontology, which indicates that hierarchical information and implicit semantic descriptions contain considerably representative and critical features for ontology matching.

## 5.5 Discussion

### 5.5.1 Impact of Ontology Enriching

According to the intuition that there are some hidden informative semantics in ontologies, especially for the complex one, we propose to enrich the ontology through ontology preprocessing and complementing. Through the statistics described in **Table 3**, numerous relationship descriptions are implicit but express a well-established role in ontology matching. Particularly in SNOMED, there are nearly more than twice the hierarchical relationships after enriching. Through the comparison of matching performance between BioHAN and

BioHAN (w/o OB) shown in **Table 2**, we can draw the conclusion that the enriching phase indeed contributes to ontology matching with the sufficient complements of semantic and structural information.

### 5.5.2 Performance Analysis of BioHAN

BioHAN uses the hyperbolic space projection to solve the intrinsical limitation in encoding complex patterns by its polynomial expanding capacity. In addition, it captures the structure of the concept by iteratively aggregating multi-hop neighborhoods with an attention mechanism. To gain an in-depth analysis of these components, we further design three variants of BioHAN: BioHAN (w/o HB), BioHAN (w/o AM), and BioHAN (w/o MN). BioHAN (w/o HB) replaces the hyperbolic projection with Euclidean space projection. BioHAN (w/o AM) removes the attention mechanism and regards all the neighboring nodes sharing the same weight. BioHAN (w/o MN) only considers the direct local neighbors and removes the multi-hop aggregation module in BioHAN. From the matching results reported in **Table 2**, we observe that the full model BioHAN achieves the best performance across all three matching tasks. It is also worth noting that both BioHAN (w/o AM) and BioHAN (w/o MN) perform better than BioHAN (w/o HB), which indicates that the hierarchical structure of the ontology captures much more essential and representative semantics. The hyperbolic graph convolutional layers can effectively encode such semantic information. By comparing the results of BioHAN (w/o AM) and BioHAN, it is obvious that the attention mechanism plays a significant role in solving the hierarchical heterogeneity of ontologies, which has improved the matching performance of 6.7% in F1 on average. For the multi-hop aggregation, by contrasting the performances of BioHAN (w/o MN) and BioHAN, it also exerts an important influence on capturing the semantics much more precisely than the complex hierarchical structures of biomedical ontologies. Multi-hop neighboring aggregation can discover much more complex matching pairs and has further improved the matching performance, especially in the measure R with an increase of 5.8% on average.

## 6 RELATED WORK

### 6.1 Biomedical Ontology Matching

Traditional feature-based approaches have been investigated for ontology matching, using terminological, structural, and semantic features for the discovery of semantically similar elements. LogMap (Jiménez-Ruiz and Cuenca Grau, 2011) uses lexical and structural indexes to enhance its scalability. AML (Faria et al., 2013) also uses various informative features and domain-specific thesauri to complete ontology matching. Feature-based matching systems mainly rely on hand-crafted features to achieve specific tasks. Unfortunately, these methods will be limited for a given scenario with weak informativeness. Representation learning has an important impact on ontology matching. OntoEmma (Wang L et al., 2018) proposes a novel

neural architecture for biomedical ontology matching, feeding into amounts of definitions and contexts to encode the concepts. It derives a variety of labeled data for supervised training and augments entities with complementary descriptions from external biomedical thesauri to improve the quality of alignments. MultiOM (Li et al., 2019) models features in ontologies from multiple views: lexical, structural, and resource. Then, it optimizes the vectors by limiting the sampling scope via structural relations in ontologies. Wang et al. (2021) systematically analyze and verify the effectiveness of multi-dimensions matching clues, subsequently aggregating the representation learning clues to boost biomedical ontology matching.

## 6.2 Graph Representation Learning

Recently, graph representation learning has gained great attention as graph neural networks (GNNs) have achieved state-of-the-art performance in various fields, such as community detection (Gargi et al., 2011), link prediction (Liben-Nowell and Kleinberg, 2007), graph alignment (Sun et al., 2018), and node classification (Bhagat et al., 2011). Some studies (Chen et al., 2017; Wang L et al., 2018) have used GNNs to achieve graph-embedded entity alignment, as similar entities often have similar neighborhoods in knowledge graphs (KG). Considering the attention mechanism, a graph attention network (Veličković et al., 2018) is proposed to learn the relatedness and importance propagated from the neighbors to the centered node. Then the neighboring message is incorporated with the measured weights. DAEOM (Wu et al., 2020) develops Siamese graph attention mechanism-based autoencoders to effectively integrate both the network structure and terminological description for deep latent representation learning in ontology matching. Recently, some researchers have substantiated that data in the form of graphs exhibit a non-Euclidean latent anatomy (Wilson et al., 2014; Bronstein et al., 2017). In addition, some recent works (Bronstein et al., 2017; Nickel and Kiela, 2017) have demonstrated the distinguished representation ability of hyperbolic manifold to model datasets with hierarchical layouts, as the hyperbolic geometry performs well in reflecting hierarchical representations naturally. Inspired by this insight, numerous research studies focus on investigating the hyperbolic geometric graph models, such as those by Nickel and Kiela (2017); Nickel and Kiela (2018); Ganea et al. (2018); and Hao et al. (2021). MEDTO (Hao et al., 2021) encodes the hierarchical features of concepts through hyperbolic graph convolution layers and further captures both local and global structural information of concepts via heterogeneous graph layers to learn better concept representations for ontology matching, and has achieved remarkably competitive performance.

## 7 CONCLUSION

In this study, we propose BioHAN for biomedical ontology matching, a hybrid graph neural network-based auto encoder to effectively integrate hierarchical structures for latent representation learning in biomedical ontology matching.

The proposed framework BioHAN executes ontology enriching to refine and complement the semantic information and hierarchical structures. Then it encodes the geometrical properties of concepts into a hyperbolic space to capture the hierarchical information through hyperbolic graph attention layers. We further implement multi-hop neighboring aggregation to incorporate both the local and global hierarchical structures with an attention mechanism to learn better concept representations for ontology matching. Our experiments conducted on a variety of biomedical ontologies demonstrate the effectiveness of BioHAN. Nonetheless, our approach only considers the *subClassOf* relationship in the ontology, which would restrict the capability of graph representation learning. In the future, it is promising to investigate some other types of non-isomorphism relations and incorporate the heterogeneous features into biomedical ontology matching. In addition, as for the large-scale biomedical ontology, the matching efficiency would also be taken into account in future research.

## REFERENCES

- Alsentzer, E., Murphy, J. R., Boag, W., Weng, W.-H., Jin, D., Naumann, T., et al. (2019). "Publicly Available Clinical BERT Embeddings," in Proceedings of the 2nd Clinical Natural Language Processing Workshop, June 2019 (Minneapolis, MN: Association for Computational Linguistics), 72–78.
- Babalou, S., Kargar, M. J., and Davarpanah, S. H. (2016). "Large-scale Ontology Matching: A Review of the Literature," in 2016 Second International Conference on Web Research (ICWR), Tehran, Iran, April 27–28, 2016, 158–165. doi:10.1109/ICWR.2016.7498461
- Bhagat, S., Cormode, G., and Muthukrishnan, S. (2011). Node Classification in Social Networks. *Soc. Netw. Data Anal.* 5, 115–148. doi:10.1007/978-1-4419-8462-3\_5
- Bodenreider, O. (2004). The Unified Medical Language System (Umls): Integrating Biomedical Terminology. *Nucleic acids Res.* 32, 267D–270D. doi:10.1093/nar/gkh061
- Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., and Vandergheynst, P. (2017). Geometric Deep Learning: Going beyond Euclidean Data. *IEEE Signal Process. Mag.* 34, 18–42. doi:10.1109/msp.2017.2693418
- Chauhan, A., Vijayakumar, V., and Sliman, L. (2018). Ontology Matching Techniques: A Gold Standard Model. *arXiv*. [Preprint]. Available at: <https://doi.org/10.48550/arXiv.1811.10191>.
- Chen, M., Tian, Y., Yang, M., and Zaniolo, C. (2017). "Multilingual Knowledge Graph Embeddings for Cross-Lingual Knowledge Alignment," in IJCAI'17: Proceedings of the 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia, August 19–25, 2017 (AAAI Press), 1511–1517. doi:10.24963/ijcai.2017/209
- Cimino, J. J., and Zhu, X. (2006). The Practical Impact of Ontologies on Biomedical Informatics. *Yearb. Med. Inf.* 15, 124–135. doi:10.1055/s-0038-1638470
- De Potter, P., Cools, H., Depraetere, K., Mels, G., Debevere, P., De Roo, J., et al. (2012). Semantic Patient Information Aggregation and Medicinal Decision Support. *Comput. methods programs Biomed.* 108, 724–735. doi:10.1016/j.cmpb.2012.04.002
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding," in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (Minneapolis, MN: Association for Computational Linguistics), 4171–4186.
- Djeddi, W. E., and Khadir, M. T. (2014). "A Novel Approach Using Context-Based Measure for Matching Large Scale Ontologies," in International Conference on Data Warehousing and Knowledge Discovery, Munich, Germany, September 1–5, 2014, 320–331. doi:10.1007/978-3-319-10160-6\_29
- Donnelly, K. (2006). Snomed-ct: The Advanced Terminology and Coding System for Ehealth. *Stud. Health Technol. Inf.* 121, 279–290.
- Euzenat, J., and Shvaiko, P. (2007). *Ontology Matching*. Heidelberg, Germany: Springer. doi:10.1007/978-3-540-49612-0
- Faria, D., Pesquita, C., Santos, E., Palmonari, M., Cruz, I. F., and Couto, F. M. (2013). "The Agreementmakerlight Ontology Matching System," in OTM Confederated International Conferences" On the Move to Meaningful Internet Systems", Graz, Austria, September 13–9, 2013, 527–541. doi:10.1007/978-3-642-41030-7\_38
- Gale, D., and Shapley, L. S. (1962). College Admissions and the Stability of Marriage. *Am. Math. Mon.* 69, 9–15. doi:10.1080/00029890.1962.11989827
- Ganea, O., Bécigneul, G., and Hofmann, T. (2018). "Hyperbolic Entailment Cones for Learning Hierarchical Embeddings," in International Conference on Machine Learning, Stockholm, Sweden, July 10–15, 2018 (PMLR), 1646–1655.
- Gargi, U., Lu, W., Mirrokni, V., and Yoon, S. (2011). "Large-scale Community Detection on YouTube for Topic Discovery and Exploration," in Fifth International AAAI Conference on Weblogs and Social Media, Barcelona, Spain, July 17–21, 2011, 486–489.
- Golbeck, J., Fragoso, G., Hartel, F., Hendler, J., Oberthaler, J., and Parsia, B. (2003). The National Cancer Institute's Thesaurus and Ontology. *SSRN J.* 1 (1), 75–80. doi:10.2139/ssrn.3199007
- Hamilton, W. L., Ying, R., and Leskovec, J. (2017). "Inductive Representation Learning on Large Graphs," in NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems (Red Hook, NY: Curran Associates Inc), 1025–1035.
- Hao, J., Lei, C., Efthymiou, V., Quamar, A., Özcan, F., Sun, Y., et al. (2021). "Medto: Medical Data to Ontology Matching Using Hybrid Graph Neural Networks," in Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, Washington, DC, August 14–18, 2021 (Association for Computing Machinery), 2946–2954. doi:10.1145/3447548.3467138
- Isern, D., Sánchez, D., and Moreno, A. (2012). Ontology-driven Execution of Clinical Guidelines. *Comput. methods programs Biomed.* 107, 122–139. doi:10.1016/j.cmpb.2011.06.006
- Jiménez-Ruiz, E., and Cuenca Grau, B. (2011). "Logmap: Logic-Based and Scalable Ontology Matching," in International Semantic Web Conference, Bonn, Germany, October 23–27, 2011, 273–288. doi:10.1007/978-3-642-25073-6\_18
- Kingma, D. P., and Ba, J. (2014). Adam: A Method for Stochastic Optimization. *arXiv*. [Preprint]. Available at: <https://arxiv.org/abs/1412.6980>

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

PW and YH outlined and wrote the overall manuscript. YH conducted the experiments. All authors actively participated in editing of the manuscript.

## FUNDING

The work is supported by the National Key R&D Program of China (2018YFD1100302) and the All-Army Common Information System Equipment Pre-Research Project (Nos. 31514020501 and 31514020503).

- Kipf, T. N., and Welling, M. (2016). "Semi-supervised Classification with Graph Convolutional Networks," in International Conference on Learning Representations (ICLR 2017) (Toulon, France: OpenReview.net).
- Kolyvakis, P., Kalousis, A., and Kiritsis, D. (2018). "Deepalignment: Unsupervised Ontology Matching with Refined Word Vectors," in Proceedings of NAACL, New Orleans, LA, June 1–6, 2018, 787–798. doi:10.18653/v1/N18-1072
- Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., et al. (2020). Biobert: a Pre-trained Biomedical Language Representation Model for Biomedical Text Mining. *Bioinformatics* 36, 1234–1240. doi:10.1093/bioinformatics/btz682
- Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N., et al. (2015). DBpedia - A Large-Scale, Multilingual Knowledge Base Extracted from Wikipedia. *Semantic web* 6, 167–195. doi:10.3233/sw-140134
- Li, W., Duan, X., Wang, M., Zhang, X., and Qi, G. (2019). Multi-view Embedding for Biomedical Ontology Matching. *OM@ISWC* 2536, 13–24.
- Liben-Nowell, D., and Kleinberg, J. (2007). The Link-Prediction Problem for Social Networks. *J. Am. Soc. Inf. Sci.* 58, 1019–1031. doi:10.1002/asi.20591
- Mungall, C. J., Torniai, C., Gkoutos, G. V., Lewis, S. E., and Haendel, M. A. (2012). Uberon, an Integrative Multi-Species Anatomy Ontology. *Genome Biol.* 13, R5–R20. doi:10.1186/gb-2012-13-1-r5
- Nejhadi, A. H., Shadgar, B., and Osareh, A. (2011). Ontology Alignment Using Machine Learning Techniques. *AIRCC's Int. J. Comput. Sci. Inf. Technol.* 3, 139–150. doi:10.5121/ijcst.2011.3210
- Nickel, M., and Kiela, D. (2017). "Poincaré Embeddings for Learning Hierarchical Representations," in *Advances in Neural Information Processing Systems* (Curran Associates, Inc.) Vol. 30.
- Nickel, M., and Kiela, D. (2018). "Learning Continuous Hierarchies in the Lorentz Model of Hyperbolic Geometry," in International Conference on Machine Learning, Stockholm, Sweden, June 10–15, 2018 (PMLR), 3779–3788.
- Otero-Cerdeira, L., Rodríguez-Martínez, F. J., and Gómez-Rodríguez, A. (2015). Ontology Matching: A Literature Review. *Expert Syst. Appl.* 42, 949–971. doi:10.1016/j.eswa.2014.08.032
- Peng, Y., Yan, S., and Lu, Z. (2019). "Transfer Learning in Biomedical Natural Language Processing: An Evaluation of BERT and ELMo on Ten Benchmarking Datasets," in Proceedings of the 18th BioNLP Workshop and Shared Task (Florence, Italy: Association for Computational Linguistics), 58–65.
- Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., et al. (2018). Deep Contextualized Word Representations. New Orleans, LA: NAACL, 2227–2237. doi:10.18653/v1/N18-1202
- Ramis, B., Gonzalez, L., Iarroyo, S., Lobov, A., Martinez Lastra, J. L., Vyatkin, V., et al. (2014). "Knowledge-based Web Service Integration for Industrial Automation," in 2014 12th IEEE International Conference on Industrial Informatics, Porto Alegre RS, Brazil, July 27–30, 2014, 733–739. doi:10.1109/INDIN.2014.6945604
- Rosse, C., and Mejino, J. L. V., Jr (2003). A Reference Ontology for Biomedical Informatics: the Foundational Model of Anatomy. *J. Biomed. Inf.* 36, 478–500. doi:10.1016/j.jbi.2003.11.007
- Schneider, T., and Šimkus, M. (2020). Ontologies and Data Management: a Brief Survey. *Künstl. Intell.* 34, 329–353. doi:10.1007/s13218-020-00686-3
- Shvaiko, P., and Euzenat, J. (2013). Ontology Matching: State of the Art and Future Challenges. *IEEE Trans. Knowl. Data Eng.* 25, 158–176. doi:10.1109/tkde.2011.253
- Suchanek, F. M., Kasneci, G., and Weikum, G. (2007). "Yago: a Core of Semantic Knowledge," in Proceedings of the 16th international conference on World Wide Web, Banff, AB, Canada, May 8–12, 2007, 697–706. doi:10.1145/1242572.1242667
- Sun, Z., Hu, W., Zhang, Q., and Qu, Y. (2018). Bootstrapping Entity Alignment with Knowledge Graph Embedding. *Proc. Twenty-Seventh Int. Jt. Conf. Artif. Intell.* 18, 4396–4402. doi:10.24963/ijcai.2018/611
- Sun, Z., Wang, C., Hu, W., Chen, M., Dai, J., Zhang, W., et al. (2020). Knowledge Graph Alignment Network with Gated Multi-Hop Neighborhood Aggregation. *Proc. AAAI Conf. Artif. Intell.* 34, 222–229. doi:10.1609/aaai.v34i01.5354
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. (2018). "Graph Attention Networks," in International Conference on Learning Representations (Vancouver, BC, Canada: OpenReview.net).
- Vrandečić, D., and Krötzsch, M. (2014). Wikidata: a Free Collaborative Knowledgebase. *Commun. ACM* 57, 78–85. doi:10.1145/2629489
- Wang, P., Hu, Y., Bai, S., and Zou, S. (2021). Matching Biomedical Ontologies: Construction of Matching Clues and Systematic Evaluation of Different Combinations of Matchers. *JMIR Med. Inf.* 9, e28212. doi:10.2196/28212
- Wang, L., Bhagavatula, C., Neumann, M., Lo, K., Wilhelm, C., and Ammar, W. (2018). "Ontology Alignment in the Biomedical Domain Using Entity Definitions and Context," in Proceedings of the BioNLP 2018 Workshop (Mwlbourne, Australia: Association for Computational Linguistics), 47–55.
- Wang, Z., Lv, Q., Lan, X., and Zhang, Y. (2018). "Cross-lingual Knowledge Graph Alignment via Graph Convolutional Networks," in Proceedings of the 2018 conference on empirical methods in natural language processing, Brussels, Belgium, October 31–November 4, 2018, 349–357. doi:10.18653/v1/d18-1032
- Wilson, R. C., Hancock, E. R., Pekalska, E., and Duin, R. P. W. (2014). Spherical and Hyperbolic Embeddings of Data. *IEEE Trans. Pattern Anal. Mach. Intell.* 36, 2255–2269. doi:10.1109/tpami.2014.2316836
- Wu, Y., Liu, X., Feng, Y., Wang, Z., Yan, R., and Zhao, D. (2019). "Relation-aware Entity Alignment for Heterogeneous Knowledge Graphs," in Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, Macao, China, August 10–16, 2019, 5278–5284. doi:10.24963/ijcai.2019/733
- Wu, J., Lv, J., Guo, H., and Ma, S. (2020). Daeom: A Deep Attentional Embedding Approach for Biomedical Ontology Matching. *Appl. Sci.* 10, 7909. doi:10.3390/app10217909
- Xie, C., Chekol, M. W., Spahiu, B., and Cai, H. (2016). "Leveraging Structural Information in Ontology Matching," in 2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA), Crans-Montana, Switzerland, March 23–25, 2016, 1108–1115. doi:10.1109/aina.2016.64
- Xue, X. (2020). A Compact Firefly Algorithm for Matching Biomedical Ontologies. *Knowl. Inf. Syst.* 62, 2855–2871. doi:10.1007/s10115-020-01443-6
- Ying, Z., You, J., Morris, C., Ren, X., Hamilton, W., and Leskovec, J. (2018). Hierarchical Graph Representation Learning with Differentiable Pooling. *Adv. neural Inf. Process. Syst.* 31, 4805–4815. doi:10.1145/3469877.3495645
- Zhang, Y., Wang, X., Shi, C., Jiang, X., and Ye, Y. F. (2021). Hyperbolic Graph Attention Network. *IEEE Trans. Big Data* 8, 1. doi:10.1109/tbdata.2021.3081431
- Zhao, M., Zhang, S., Li, W., and Chen, G. (2018). Matching Biomedical Ontologies Based on Formal Concept Analysis. *J. Biomed. Semant.* 9, 1–27. doi:10.1186/s13326-018-0178-9
- Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., et al. (2020). Graph Neural Networks: A Review of Methods and Applications. *AI Open* 1, 57–81. doi:10.1016/j.aiopen.2021.01.001

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Wang and Hu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.