[1,*]Haobo Mao

Research on the Design of Teaching Aid Resource System Based on Unsupervised Representation Learning

Journal of Electrical Systems

Abstract: - In the absence of manual annotation information, using only the topology information of the graph to achieve unsupervised graph alignment has always been one of the important challenges faced by graph data mining, especially in the large-scale graph alignment task, the discovery of initial seed nodes and the low computational efficiency have always been a problem. This paper proposes a large-scale unsupervised graph-aligned physical education teaching resource system framework based on topological representation learning, which is highly scalable and adaptable. Firstly, representative subgraphs are selected from the graph to be matched as seed node candidate sets, and the local topology information is used to calculate to obtain highly reliable seed node matching results; then, the obtained seed nodes are used to integrate the graph to be matched, and an efficient unsupervised representation learning algorithm is proposed to map the fusion graph to a unified vector space; finally the obtained node vectors are used to achieve the alignment of large-scale graphs. The proposed technique shows the shortest time when dealing with large-scale graph alignment tasks, and at the same time reaches the highest level of accuracy of alignment results, in addition, the performance of the physical education auxiliary resource system algorithm has the least impact on the difference of graph structure.

Keywords: Graph Alignment, Large-scale networks, Unsupervised Means Learning, Fusion Map Mapping, Physical Education Auxiliary Resource System.

I. INTRODUCTION

In many areas of physical education support resource systems, graph data structures are an important form of modeling, which can be used to represent relationships between nodes, such as social networks in online social media and collaborator networks in scientific research, and protein networks in the field of bioinformatics [1]. Through graph data mining, we can obtain a lot of valuable information, such as node classification can infer the research field to which researchers belong, node clustering can analyze the interest community in social networks, link prediction can recommend more friend relationships for users, node importance assessment can analyze the importance of protein in biological functions, and the influence of users in the social teaching auxiliary resource system.

In recent years, we have worked on graph alignment tasks to identify whether nodes in different graphs belong to the same entity, thereby improving the recommendation accuracy of the system. In view of the important application value of graph alignment tasks, a variety of graph alignment algorithms have been proposed, one of which is the traditional discrete graph alignment algorithm, such as BigAlign2 [2]. The algorithm transforms the graph alignment problem into the optimal solution problem of the permutation matrix, and realizes the alignment of the protein network through the similarity of the node neighborhood topology; the second is the graph alignment algorithm that combines two types of technologies, such as graph feature extraction and classification method and graph matching technology. Final proposes three metrics, namely structural similarity, node feature similarity, and edge feature similarity, to solve the graph alignment problem, while the second category is alignment methods based on graph representation learning, such as regale [2].

Firstly, the method learns the vector representation of nodes by calculating the similarity of neighbor nodes, and uses the similarity of node vectors to complete the graph alignment task; secondly, the label propagation technology is introduced in the graph representation learning process to reduce the number of labeled samples required for graph alignment. CrossMNA [3] adopts an unsupervised graph alignment framework based on multilayer vector representation to achieve node alignment by optimizing two types of node vectors and using the similarity of vectors between graphs, while Galign [4] uses a graph convolutional network (GCN) to learn the vector representation of nodes and proposes a data augmentation method to ensure that the model satisfies the consistency constraints of the graph. In the image matching of the physical education auxiliary resource system, if the lighter level GCN is used combined with the image representation method in the initial state, it may face some challenges, such as when dealing with nodes with similar distribution probabilities, it is difficult to find the optimal node correspondence due to the complex association relationship between the nodes. In addition,

[1] JINHUA POLYTECHNIC, Zhejiang, China
*Corresponding author: Haobo Mao

traditional graph alignment algorithms are usually only suitable for image data in a single domain and lack versatility, while IsoRank [5] algorithms are a method specifically designed to deal with protein network alignment problems.

In this paper, an efficient graph representation learning method combining sparse matrix decomposition and personalized random walk is proposed, aiming to maximize the capture of global structure information while ensuring computational efficiency. To verify the effectiveness of the algorithm, we performed experimental validation on six datasets of different types and scales and compared it with multiple known state-of-the-art unsupervised graph alignment algorithms. Experimental results show that the proposed method can quickly and efficiently achieve the matching alignment of large-scale graphs, and achieve the highest matching accuracy. Even when the structure of the two matching plots is quite different, we can still maintain a high matching accuracy. In addition, this paper also introduces label propagation technology to improve the prediction performance of the model, obtain better classification results, and further improve the robustness of the physical education auxiliary resource system [6]. The main contribution of this paper is to propose a large-scale graph alignment algorithm based on seed node fusion, which is different from the previous method of using seed node supervised graph representation learning, which directly maps the graph data to be matched into a unified vector space by performing unsupervised graph representation learning on the fusion graph, thereby avoiding the problem of sharp decline in matching accuracy due to representation space transformation. By analyzing the experimental results of a large number of public datasets, it is shown that the algorithm can not only quickly and accurately complete the automatic registration between large-scale graph pairs, but also obtain good robustness and fault tolerance. We propose a graph representation learning method combining sparse matrix decomposition and personalized random walk, which can efficiently learn large-scale graphs and obtain representation vectors that can capture the global structure information of the graph, so it is very suitable for graph alignment tasks [7]. Aiming at the challenges faced by large-scale graph alignment, such as the feature dimension is too high and it is difficult to obtain effective feature subsets, a rapid optimization strategy for massive graph data is proposed. After extensive experimental verification, the proposed method shows excellent matching accuracy and computational efficiency of the physical education auxiliary resource system when processing large-scale graph data, which far exceeds other methods. In addition, aiming at the problem of a large number of redundant features in large-scale graph model training, a hybrid deep neural network framework for graph classification and image retrieval is designed.

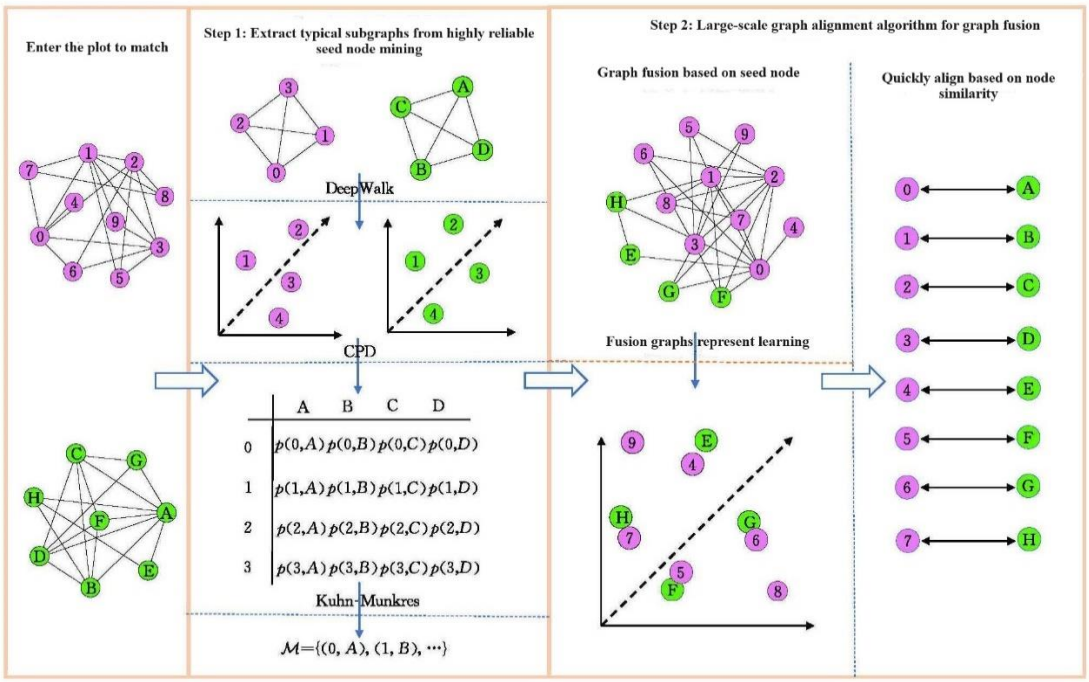

Figure 1: Schematic diagram of LGA large-scale diagram alignment frame

II. RELATED WORK

A. *Conventional Graph Alignment Algorithms*

Graph alignment has important applications in several areas of teaching aid resource systems. For example, in the field of biology, we can study the relationship between proteins, while in the field of physical education

support resource system, we can use cross-platform physical education auxiliary resource system network user alignment for user profiling and user interest mining. In the field of knowledge graph, we can complete and expand the knowledge graph. In the traditional graph alignment algorithm, we usually directly calculate the alignment matrix S and input the source and target graphs in the form of adjacency matrices. During the alignment process, we continuously update the alignment matrix by optimizing the loss function, so that the adjacent matrix of the arranged target graph is close to the adjacency matrix of the source graph. In order to solve the existing NP problem, we can find the permutation matrix H by optimizing the loss function, thereby improving the efficiency of the algorithm [8].

B. Graph Alignment Methods Based on Graph Representation Learning

With the continuous evolution of graph representation learning technology, node-based graph alignment algorithms are gradually becoming a hot topic in academic research on teaching auxiliary resource systems. Due to the shortcomings of traditional graph representation learning algorithms, such as poor scalability, difficulty in processing heterogeneous data, and inability to effectively mine hidden relationships, this paper analyzes and summarizes the relevant research results in the field of physical education learning represented by graph representations and proposes improvement schemes. When studying the social network user alignment problem, IONE [10] proposes a modeling method that takes each user's followers (follower-ship) as a contextual representation of input/output to keep the representation vectors of users with similar followers/followers closer together in the embedding space. In order to further solve the problem of inconsistent embedding space of nodes in different graphs, IONE selects nodes with large matching probabilities as anchor nodes, so that it provides higher weight of context embedding representation in space for neighboring nodes. Finally, the stochastic gradient descent and negative sampling algorithms are used to solve the scalability problem. REGLE obtains the initial representation vector of the node by weighting the degree information of the first-order neighbor and the second-order neighbor of the node, and calculates the similarity of the node according to the similarity of the initial representation vector, so that the dimensional graph with different structures has a higher general similarity [11].

III. HIGHLY RELIABLE SEED NODE MINING

In the node sets of G1(v1, E1,) and G2,(v2, E2,) graphs, v1 and v2 represent the sets of E1 and E2 sides, respectively. For a clearer and more convenient expression, v is used to represent the nodes in Figure G2. In this way, the two subgraphs composed of v1 and v2 have the most similar topological connection structure, and in the absence of prior information and node attribute information, the unsupervised alignment of G1 and G2 mainly depends on the topological information between them. However, in the case of large scale of the system diagram of the physical education auxiliary resource, direct matching of the whole map tends to be less efficient and has a high error rate. In order to solve the above problems, this paper proposes a similarity measurement method between any two adjacent vertices in the source graph based on the combination of graph theory and probability and statistics knowledge, and verifies the correctness and effectiveness of the algorithm by examples [11]. In order to further perform matching calculation, this paper first selects a typical subgraph as the candidate seed node set in the graph and aligns it, so as to obtain a highly reliable matching node pair between the source graph G1 and the target graph G2, which lays a solid foundation for the next matching calculation.

A. Typical Sub-graph Extraction

The degree of nodes of graph structures constructed from real-world data tends to follow a power-law distribution, that is, only a small number of nodes have a large number of neighbor nodes and most nodes have fewer connections. Figure 2 shows the distribution of the nodality of the six public datasets involved in the subsequent experiments in this paper, and it can be seen from the figure that all datasets show that a small number of nodes have a large number of connections, which contain richer structural information, and the accuracy and reliability of the results obtained by matching and aligning these nodes will be higher. Therefore, in this paper, the top K nodes with the largest degree are selected from the graph to be matched to form a typical subgraph, and then the seed nodes are obtained by unsupervised graph alignment of the typical subgraph [12]. For the matching graph G1 (v1, E1), the first K nodes with the largest degree are selected to form the set v1={u1, u2, ..., uk},Then v1′ is a subset of v1, let the subgraph composed of elements in v1′ be G1′ (v1′, E1′), where E1′ is a subset of the original edge set E1, and only retain the edges between nodes in v1, so the resulting subgraph G1′ (v1′, E1′) is a typical subset of G1(v1, E1), because the nodes in v1 are the largest in the original graph, so the typical subgraph G1′(v1′, E1′) retains most of the topology information of the original graph.

Assuming that the typical subgraphs of G2, (v2, E2,) are G2, (v2, E2,), then we can reach this conclusion. The high sparsity of high-dimensional data makes it difficult to achieve the desired effect in practice, so fewer low-dimensional edges can be introduced for processing. The highly reliable seed nodes discussed in this section are the set of node pairs containing v1′ and v2′ subsets, which are also M subsets. The initial point generation method based on genetic algorithm optimization can solve the shortcomings of traditional genetic algorithms that are easy to converge prematurely and fall into local optimal solutions to a certain extent, and is an effective genetic algorithm [13]. Under the premise of ensuring that K selection adapts to the scale of the original graph, considering that the algorithm is universal, that is, no matter how large the graph size is, the algorithm can efficiently and accurately select the appropriate number of seed nodes to achieve the optimal selection of K [14]. For larger scale graphs, selecting seed nodes by percentage may reduce seed matching accuracy and pre-matching efficiency; and for smaller graphs, there will be insufficient number of seed nodes, and it is difficult to fully play the pre-matching role. In order to solve the shortcomings of the above physical education auxiliary resource system, a method for quickly determining the number of seed nodes based on scale factor is given. The calculation formula given in this paper can avoid the problem of pre-matching too many nodes selected by large-scale networks, and can also avoid the selection of too few nodes for small graphs to construct typical subgraphs to achieve algorithm adaptation results [15-17].

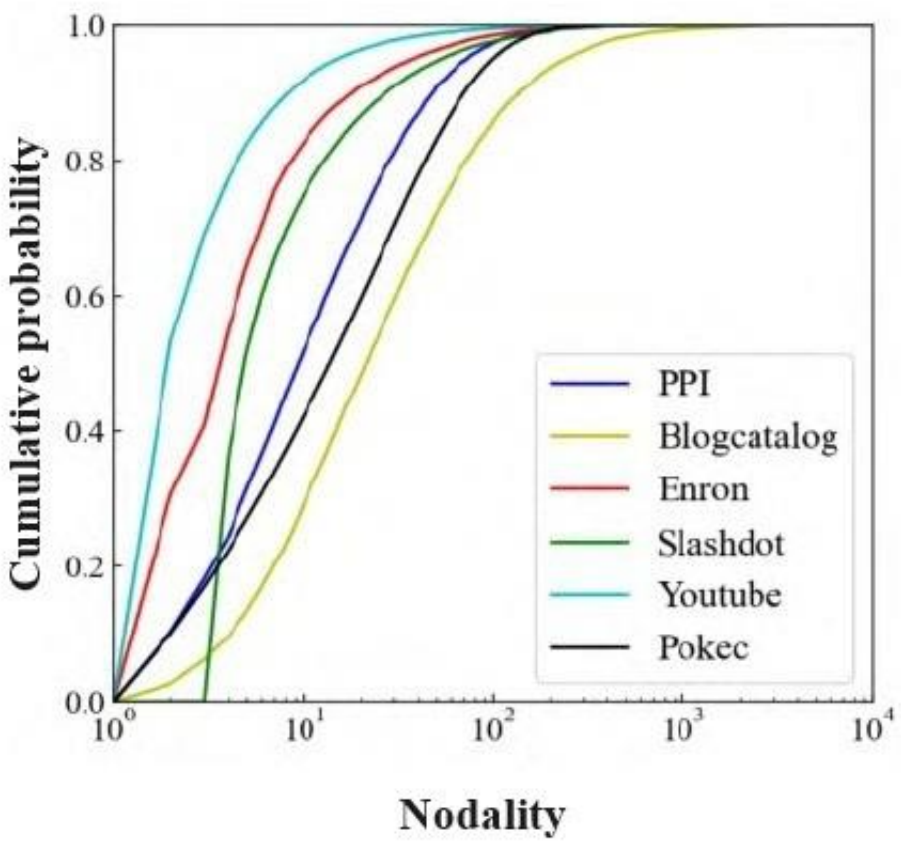

Figure 2: Cumulative probability distribution of node degree in different graphs

B. Typical Sub-graph Matching

The DeepMatching algorithm is used to perform unsupervised alignment of typical subgraphs to obtain highly reliable seed nodes, first the DeepWalk graph representation learning algorithm is used to obtain the node vector representation of the two typical subgraphs, and then the consistency point drift algorithm (CPD, Coherent Point Drift) is used to obtain the matching probability matrix between the nodes to be matched, and finally the maximum weight matching algorithm is used to obtain highly reliable seed nodes [18-23].

1) Typical Sub-graph Representation Learning: Assuming that the extracted typical subgraphs are G1′(v1′, E1′) and G2′(v2′, E2′), then we will first encode these two subgraphs using the DeepWalk[24] algorithm to obtain the vector representation of the nodesThe DeepWalk algorithm consists of two steps. The random walk algorithm is used to generate node context neighbor node information. When walking randomly, each time a node is randomly selected from the neighbors of the node as the next hop, and the cycle continues until it reaches a certain path length and stops walking. In order to avoid repeatedly traversing all possible edges and calculating their cost

function, an improved genetic algorithm with the objective function value minimization as the fitness function is designed to solve the problem. When studying the node vector representation problem, the Skip-Gram model word2vec technology is used to treat the nodes in the graph as word2vec language model words, and the path obtained by random walking is regarded as natural language sentences. The model learns node vector representations with maximized node context probabilities. Step 3, divide the network to obtain sub-atlases of different sizes, and use the weighted average method to calculate the weight values of each node in the sub-atlas to obtain the optimal weight matrix.

2) Consistency Probability Matrix Calculation: We use the consistency point drift algorithm to model the correlation probability problem between nodes in the R1′ and R2′ concentrations, and use the probability values to characterize the matching probability between nodes to judge the degree of certainty of the correspondence between nodes in the two point concentrations. Based on this result, it can be determined whether there are potential topologies at two points, and how these potential topologies affect the overall performance of the network. The mixed Gaussian Mixture Model (GMM) is used to describe the deterministic relationship between all node pairs in the point set R1′ and R2′, then the Gaussian mixture model parameters are solved to obtain the matching relationship between nodes, the maximum likelihood method is used to fit the Gaussian mixture model centroid to maximize the probability density function expectation, and finally the mixed Gaussian model parameters are solved to obtain the matching probability between nodes [25]. Due to the high accuracy of Gaussian mixture models, they can be applied to relative position estimation between unknown and known nodes for search space optimization and faster computation. In the expected solution stage (E-step), the GMM model uses the EM algorithm to calculate the matching probability matrix P between the seed nodes of the two graphs, and calculates the elements P_{ij} by the following method.

$$p_{ij} = \frac{\exp(-\frac{1}{2\sigma^2}\| r_j^{\,2} - (_{ri}^{\,2} + G_{i,w})W \|^2)}{\sum_{k-1}^{n_2} \exp(-\frac{1}{2\sigma^2}\| r_j^{\,2} - (_{ri}^{\,2} + G_{i,w})W \|^2) + \frac{w}{1-w}\frac{(2\pi\ \sigma^2)n_2}{n_1}} \tag{2}$$

Thereinto:

W is the coefficient matrix, initialized to the all-zero matrix, and σ^2 is the isotropic orthorhombic variance matrix,

The initialized values are:

$$\sigma^2 = -\frac{1}{dn_1n_2}\sum_{i,j-1}^{n_1,n_2}\| r_j^1 - r_i^2 \| \tag{3}$$

The maximization stage (M-step) mainly updates the coefficient matrix, which is calculated as follows:

$$W = (G + \lambda\sigma^2 diag(P1)^{-1})^{-1}(diag(P1)^{-1}PM_1 - M_2) \tag{4}$$

Among them: diag is diagonalization operation, P is the matching probability matrix, 1 is the column vector of all 1s, and the matching probability matrix P. After iterative convergence, the matching probability between node ui and vj can be obtained, the higher the probability value, the greater the certainty of the correspondence between the two nodes.

C. Seed Nodes Generation

The method in this paper is to establish an unsupervised graph alignment algorithm based on topological information, and most of the existing unsupervised algorithms are more or less profitable when using topological consistency constraints, the current distribution of these two graphs is quite different, in large cases, the alignment tasks are based on supervision, such as Wang and others [26]. Personnel use a double constraint mechanism to balance the two diagrams where the degree of different nodes is very different. At present, most of the unsupervised algorithms assume that the topological consistency between the two graphs after alignment is calculated when the topology of the two graphs to be matched is very different, and the accuracy of normal alignment is usually not high, and the similarity between nodes is not too high.

In order to obtain highly reliable seed node pairing, the paper examines the sub-node consistency probability matrix by species using a bipartite graph's maximum weight matching algorithm to maximize the probability of matching between seed nodes. Dichotomous graph matching aims to find a sub-graph from a bipartite graph and

request that any edge belonging to this sub-graph will not intersect at the same vertex point and the sum of the weights of all the edges of this sub-graph is maximum.

In this paper, the set of seed nodes MS is obtained by using the Kuhn-Munkres algorithm. In this paper, the K vertices with the highest degree in the 2 graphs are selected for the construction of the canonical sub-graphs because it maintains the topological information of the original graphs to the maximum extent possible, thus ensuring that the seed nodes of the obtained matches are highly reliable. It is worth mentioning that the nodes with the larger degree are not necessarily matched with the corresponding nodes in other graphs to ensure that the seed nodes are reliable, and only retain the matching probability of more than 1 fixed-gap value seed node pairing to avoid a greater negative impact on the subsequent matching.

When the two graphs are very different in topology when matching the vast majority of seed nodes match probability will be less than this 17 value, if the number of seed nodes when the node is less than 0.5K, that the source graph and the target graph topology is very different and does not perform the subsequent fusion of graph matching alignment to ensure that the final results are reliable and effective method in the paper is to build an unsupervised graph alignment based on the topological structure information. Algorithms, most of the existing unsupervised algorithms are more or less favorable when using topological consistency constraints, the current distribution for these two graphs differ greatly in large examples, the alignment task are built on a supervised basis, such as Wang and others [27] who use a double constraint mechanism to balance the two graphs in the different node degree difference is very large occasions.

Most of the current unsupervised algorithms assume that the topological consistency between the two graphs after alignment is calculated when the topologies of the two graphs to be matched are very different from each other, and the alignment accuracy is usually not high, and the similarity between the nodes will not be too high. In order to obtain highly reliable seed node pairing, the paper is studied by species sub-node consistency probability matrix using bipartite graph of maximum power matching algorithm makes the matching probability between seed nodes to reach the maximum. Dichotomous graph matching aims to find a sub-graph from a bipartite graph and request that any edge belonging to this sub-graph will not intersect at the same vertex point and the sum of the weights of all the edges of this sub-graph is maximum.

In this paper, the set of seed nodes MS is obtained using the Kuhn-Munkres algorithm. In this paper, the K vertices with the highest degree in the 2 graphs are chosen for the constructed sub-graph of the canonical model because it maximizes the preservation of the topological information of the original graphs and thus ensures that the seed nodes obtained for matching are highly reliable.

It is worth mentioning that not a larger degree of nodes will have a corresponding matching node existing in other graphs in order to ensure that the seed nodes are reliable, only keeping the matching probability of more than 1 fixed gap value of the seed node pairing to avoid a greater negative impact on the subsequent matching. When the two graphs in the topological structure of a very large difference in the matching of the vast majority of seed nodes match probability will be less than the value of 17, if the seed when the number of nodes is less than 0.5K, that the source graph and the target graph topological structure of the difference is very large and do not perform the subsequent fusion of graphs to match the alignment so as to ensure that the final results of the reliability and validity of the results obtained.

IV. Large-scale graph Alignment Algorithm Based on Graph Fusion

When the seed node set is obtained, the existing supervised graph alignment algorithm G1 can be used to match the alignment with G2. Most of the existing methods use graph representation learning algorithms to obtain the vector representations of G1 and G2 nodes separately, and then transform the G1 and G2 vector representations to the same vector space by seed nodes. The process, due to the inconsistency between the node distribution of the seed node Forestry Department and the node distribution of the whole graph, may cause the transformed vector representations to produce a large error, thus affecting the accuracy of the matching results, and because of the large scale of G1 and G2, most of the existing graph represents ∪ learning algorithms. In order to solve the above problems, the paper proposes a new large-scale graph alignment algorithm based on graph integration: first, the graphs to be matching graph is integrated into a large graph according to the sub-nodes in the seed graph, then a new effective graph representation learning algorithm is proposed to get the vector representations of the nodes, and finally the node vectors are used to compute their similarity, thus realizing the pi-class matching [28].

A. Seed Node-based Graph Fusion

In order to make the paired high-reliability seed nodes have more structural information weight in the fusion graph and reduce the negative impact caused by the overall structure difference, this paper fuses G1 and G2 based on the matched seed nodes to construct a new large graph G3. Specifically, the matching seed nodes in G1 and G2 are aggregated to form a new node, and the remaining nodes in G and G2 are added to G3 according to the connection relationship between the original nodes, this process can be regarded as the connection between the network and G2 on the basis of the seed node, and the seed node is the link of the connection.

B. Fusion Diagrams Represent Learning

After many rounds of wandering, the possibilities of each vertex in the graph to be visited by the source node, will gradually converge and stabilize, and the stabilization possibilities can be used as the source node, affinity scores with other nodes. Personalized random wandering of all the nodes in the graph individually as source nodes can yield the following affinity matrix:

$$M = \sum_{i=0}^{\infty} \alpha(1-\alpha)^i \bullet P^i \tag{5}$$

In the formula: M is a dense matrix.

Singular value decomposition of the adjacency matrix A of Figure G3:

$$A = U \sum V^T \quad \text{A} \tag{6}$$

In the formula: A is the left singular matrix and V is the right singular matrix.

In order to gradually aggregate the higher-order structural information in the graph, an iterative computation of the initial embedding vector 1 can be obtained:

$$x = \sum_{i=1}^{\infty} \alpha(1-\alpha)^i \bullet P^{i-1} X_1 .\#(4) \tag{7}$$

The iterated vector representation α contains higher-order structural information in the graph, of which the value of the parameter α determines the range of the graph structural information contained in the representation vector, if the value is too low, the representation vector α can contain more higher-order information, but it needs more iterations to converge; if the value is too low, it will contain more local structural information, for example, if the figure α is 1 in the extreme case, then the obtained representation vector contains only the first-order neighborhood structural information of the node, which is the first-order neighborhood structural information. For example, if the value is 1 in the extreme case, the learned representation vector contains only the first-order neighbor structure information of the node. In order to synthesize the effectiveness and efficiency of the algorithm, the figure α of 0.15 is used in this paper.

C. Rapid Dissemination of Higher-Order Information Based on Community Structure

The computation and aggregation of higher-order information in the teaching aid resource system requires a large number of iterative operations of large-scale matrix multiplication, thus resulting in old low computational efficiency [29]. In order to improve the efficiency of this part of the computation, the paper adopts the widely existing community structure in the graph to accelerate the propagation process of higher-order information. The graphs formed in the real world tend to have a certain degree of aggregation phenomenon, also known as community structure, where the connectivity between nodes in the same community is relatively tight, while the connectivity between nodes that belong to different communities is relatively sparse, and the community structure widely exists in real graph data, for example, the connectivity between users with the same interest in a social network will be more, thus forming interest communities; and in the biomedical field, researchers often link proteins with similar biological functions to drug-use edges, which form community structures on biological protein networks, see Figure 3. Two kinds of cell structures appear in the figure, and it is obvious that the nodes within a cell are more closely connected than the nodes between cells. The diffusion of information depends on the node connections, and the denseness of the node connections will have a great impact on the diffusion efficiency, so the diffusion of the node information within and between the cells is very different, and in the paper, we use the characteristic of the figure with the community structure to accelerate the higher-order information diffusion computation [30].

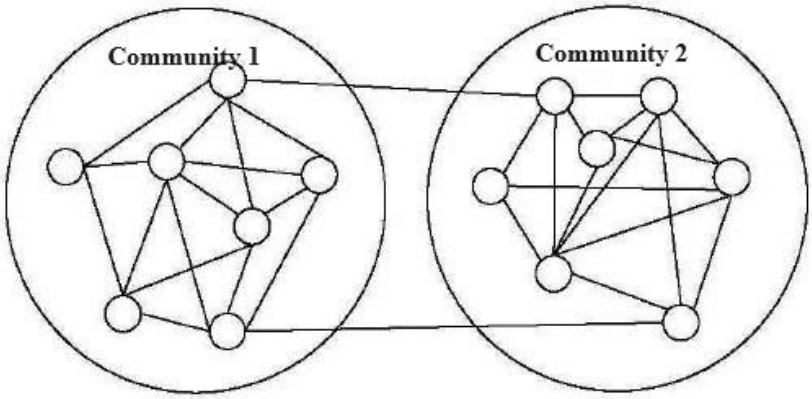

Figure 3: The phenomenon of community aggregation

In order to obtain the representation vector I of the graph, the process can be seen as the continuous propagation and aggregation of information on the graph. Due to the existence of the community structure in the graph, the nodes in the community are closely connected, resulting in repeated dissemination of node information in its corresponding community, and it is difficult to spread outside the community, and only after a certain number of iterative propagation can the node information gradually spread to the whole map; the high-order information in this graph is integrated into the initial, representing the vector is not neutral, and the process of collecting and disseminating this information is very cumbersome and difficult to complete, so we have to constantly collect and integrate information. At the same time, we also need to note that due to the existence of social E structure, the connection between nodes in the community is very close, which also leads to repeated dissemination of node information in the corresponding community, this process is very difficult, and even not easy to be leaked by information outside the community, can only be completed through continuous efforts. Therefore, the propagation process of this information in the figure can be divided into 2 stages: community propagation and escape propagation, and based on the discovery, the representation vector X can be quickly calculated.

Escape Propagation Stage: In this phase, the propagation of information will no longer be limited by the community structure in the diagram, and the closer to the source node, the more information the node will obtain from the source node, a process known as the escape phase. Therefore, through the degree of calculation, it can be found that the closer to the center of the propagation source, the closer the relationship between the points in the network becomes. In the escape stage, after many iterations, the correlation between the information and the damage of the source node is continuously reduced after the nodes in the graph are propagated at a higher level. At this time, the correlation between the information and the source node after the high-order propagation of the node in the graph is continuously reduced, that is, the information obtained by the node in the escape stage has nothing to do with the distance to the propagation source, and the obtained damage information is closely related to the information owned by the neighbor node, using the original PageRank algorithm, the information propagation in the escape stage can be accurately calculated to avoid the occurrence of missing information. On this basis, an analysis method for competition and cooperation between nodes in multi-hop networks based on the combination of graph theory and evolutionary game theory is proposed, and the feasibility of the proposed method is verified by simulation experiments.

D. Fast Matching Alignment Based on Node Similarity

When the nodes obtain vector representations, this paper adopts a fast matching multi-alignment algorithm to solve this problem. The algorithm makes use of the special network structure composed of a large number of points with the same structure and different topological properties in the graph and network, which effectively reduces the operation time. Firstly, the vector Euclidean distance calculation is carried out to describe the spatial similarity of nodes, and the core algorithm of this paper refers to the intention in the figure when dealing with the problem, but in fact indirectly borrows the intention therein.

In this paper, a large-scale graph alignment technique for physical education auxiliary resource system based on seed graph alignment node integration is proposed, which is different from the traditional graph alignment learning method based on graph representation, which avoids the graph alignment transformation in the representation space by performing unsupervised graph representation learning on the integrated graph, thereby significantly improving the matching accuracy. In addition, by reconstructing the original network model and merge matrix of the physical education auxiliary resource system, the robustness and noise immunity of the

network model are further enhanced. In order to improve the computational efficiency of the large-scale fusion graph representation learning algorithm, the algorithm adopts the community structure of the graph to accelerate the calculation speed of high-order structure information to the graph representation learning, so as to greatly improve the efficiency of graph representation learning under the premise of ensuring effectiveness, so that the algorithm in the paper can effectively solve the problem of large-scale graph representation alignment.

V. EXPERIMENTS AND RESULTS ANALYSIS

A. Experimental Data and Operating Environment

On six widely used public datasets, we conducted experiments on public datasets, the Teaching Aid Resource System experiment, and detailed information about this dataset in Table 1. All datasets are available on public websites or papers, which is a source of information that is not left out. This paper focuses on solving the problem of unsupervised graph alignment based on topology information, where the dependencies of the algorithm are limited to the topology of the graph, and since the two graphs to be matched are usually corrupted, they tend to share a common connection relationship. For example, the WeChat social network and the QQ friend relationship network are based on the real acquaintance relationship between users, but due to the randomness of human activities and incomplete data collection, there will be a certain degree of damage and difference in the topology of the two social networks, but they can be regarded as obtained from the user's real social network according to a certain proportion of sampling connection relationship. To ensure the stability of the experimental results, we need to repeat the execution 20 times to obtain more accurate results, while constantly cropping and generating plots to be matched to meet the uncertainty requirements. From the results, random clipping did not cause significant damage to the running results, so the operation has a wide range of applicability. In all experiments, only the graph structure information is used to complete the alignment task, and no corrupt matching information is found initially, which requires special attention.

Table 1: Experimental data set

THE DATASET NAME			
PPI	TYPE	NUMBER OF NODES	NUMBER OF EDGES
BLOGCATALG	Protein networks	3890	76584
ENRON	Social networks	95	10312
SLASHDOT	Social networks	36 692	183831
YOUTUBE	Mail communication network	77 360	905468
POKEC	News Network	1134890	2987624

1. PPI (Protein-Protein Interaction):

Type: Protein network

Number of nodes:3890

Number of sides:76584

Analysis: The PPI dataset represents the interactive relationships between proteins, which are important in bio-informatics and molecular biology. It can be used in the Teaching Aid Resource System to teach courses related to bio-informatics, biological network analysis and molecular biology. Students can use this dataset for experiments and research to understand the interaction networks between proteins.

2. Blog Catalog:

Type: Social Networking

Number of nodes:95

Number of sides:10312

Analysis: The Blog Catalog dataset contains social relationships between bloggers and is suitable for studying social network analysis and users' behavior in social media. It can be used in the teaching resource system to introduce the basic concepts of social network analysis and social media marketing, as well as for practical data analysis and experiments.

3. Enron:

Type: Social Network

Number of nodes:36,692

Number of sides:183,831

Analysis: The Enron dataset is based on email communications from Enron and can be used to study social network analysis, email communication patterns, and data privacy. It can be used in the Teaching Aid Resource

System for courses on network analysis, data mining and information security, as well as for discussions on privacy protection.

4. Slashdot:

Type: Mail communication network

Number of nodes:77,360

Number of sides:905,468

Analysis: The Slashdot dataset is based on email communication relationships between users of the Slashdot website. It can be used to study email communication networks and social network analysis. In the field of education, it can be used to teach concepts related to email communication, web analysis and social networks.

5. Youtube:

Type: News Network

Number of nodes:1,134,890

Number of sides:2,987,624

Analysis: The Youtube dataset contains the relationship between users and videos on YouTube. It can be used to study user behavior, social network analysis, and content distribution on the video sharing platform. It can be used in teaching aid resource systems for courses such as social media analysis, data mining and user behavior analysis.

6. Pokec:

Type: Social Network

Number of nodes:1,632,803

Number of sides:30,622,564

Analysis: The Pokec dataset is based on Slovakia's social network and contains social relationships between users. It is a large-scale social network dataset that can be used to study social network analysis, user behavior and social recommendation. It can be used in physical education resource systems to gain insight into the structure and properties of social networks and the principles of recommendation systems.

Based on these parameter settings, we can obtain an optimal model that has a short computation time and can meet certain accuracy requirements. In order to ensure the performance of this physical education auxiliary resource system model, we adopt a new initialization strategy to improve the computation speed, which is designed based on the random wandering theory. Experiments show that this model has high computational efficiency compared with other methods. In order to ensure that the model effect of Artemisia is not affected by different parameters, we adjusted the values of the special analysis parameters, and based on this, we carried out the artistic setting of the centralized residual parameters, and constantly pursued the default values in Table 2 to keep the same.

In order to make the results more reliable, we use genetic algorithm to adjust the relevant parts of the model appropriately. In order to optimize the data transmission efficiency of the model, we introduced the concept of "side" and combined it with the network flow technique to achieve more efficient data transmission. In order to measure the accuracy of various algorithms in graph alignment matching, we adopt the present leakage node pairs as the evaluation index, and select the node pairs with the highest matching probability as the alignment nodes in the calculation process.

Table 2: Experimental parameter setting table

PARAMETER NAME	PARAMETER DESCRIPTION	RANGE OF VALUES	DEFAULT VALUE
α	Stopping probability of personalized random wandering	0~1.0	0.15
s	Number of community dissemination iterations	6~18	9
L	Number of initial iterations of community dissemination	2~7	4
β	Adjustment factor for the number of nodes in a typical subgraph	2~12	4
D	The dimension of the embedding vector	40~18	80
0	Threshold for seed node matching probability	0~1.0	0.90

1. α (Stopping probability of personalized random wandering):

Parameter description: This parameter controls the stopping probability of the personalized random walk, i.e., in the random walk, the probability of stopping or continuing at each step.

Value range: 0 to 1.0, indicating the range of stopping probability.

Default value: 0.15, meaning that there is an 85% probability of continuing and a 15% probability of stopping during a random walk.

2. s (Number of community dissemination iterations):

Parameter description: this parameter controls the number of iterations of the community propagation algorithm, i.e., how many times the algorithm will iterate to find the community structure during the community detection process.

Range of values: between 6 and 18, indicating the range of the number of community propagation iterations.

Default value: 9, indicating that in default the algorithm will iterate 9 times to find the community structure.

3. L (Number of initial iterations of community dissemination):

Parameter description: this parameter specifies the initial number of iterations of the community propagation algorithm, usually a certain number of iterations at the beginning of the community detection, and then more iterations.

Value range: between 2 and 7, indicating the range of the initial number of iterations for community propagation.

Default value: 4, indicating that by default 4 iterations are performed at the beginning of community detection.

4. β Adjustment factor for the number of nodes in a typical sub-graph:

Parameter Description: This parameter is used to adjust the number of nodes in a typical sub-graph, which is the sub-graph used to represent data in the resource system.

Value range: between 2 and 12, indicating the range of the adjustment factor.

Default value: 4, which means that by default the adjustment factor used to adjust the number of nodes in a typical sub-graph is 4.

5. D (The dimension of the embedding vector):

Parameter description: this parameter determines the dimension of the embedding vector, i.e. the dimension of the representation of each node in the embedding space.

Value range: between 40 and 180, indicating the range of the embedding vector dimension.

Default value: 80, indicating that by default the dimension of the embedding vector for each node is 80.

6. 0 (Threshold for seed node matching probability):

Parameter description: this parameter is used to determine the probability threshold for seed node matching, i.e., the probability threshold that needs to be reached for nodes in two graphs to match successfully in graph alignment.

Value range: between 0 and 1.0, indicating the range of the matching probability threshold.

Default value: 0.90, which means that by default the two nodes need to reach 90% probability threshold for successful matching.

These parameters play a key role in the design study of teaching aid resource systems, and they can be used to adjust the performance and behavior of the system to meet different scenarios and needs. The researcher can adjust these parameters to optimize the performance and effectiveness of the system based on specific experiment and application requirements. The ranges and default values of these parameters provide flexibility and a starting point for experimentation and optimization in the study.

B. Ablation Experiment

The effectiveness of the community structure-based higher-order information propagation (LGA-Comm) mechanism in graph representation learning was verified by using ablation experiments. In the representation learning phase, we set up direct iteration for higher-order information propagation (LGA-Full) and non-information propagation (LGA-Adj), and performed and compared the graph alignment task on four datasets, namely, PPI, BlogCatalog, Enron, and Youtube.

In the experimental process of the sports teaching aid resource system, we set the proportion of removed edges to 10%, and obtained the matching accuracy in accordance with different propagation methods, and the specific results are shown in Figure. 4. The result of high correct rate for all four types of graphs illustrates that this method can effectively improve the classification accuracy of the network and has a certain degree of universality. The higher-order information propagation mechanism based on the community structure of the Physical Education Teaching Aid Resource System shows almost the same effect as the direct iterative information propagation in the graph alignment task, and its matching accuracy is significantly higher than that of the mechanism without information propagation.

In this section, we compare the efficiency of higher-order information dissemination based on community structure with direct iterative information dissemination, and the results show that information dissemination based on community structure takes only one-tenth of the time of direct iterative computation, and this method has a significant efficiency advantage, as shown in Table 3. The node representation vectors can only obtain the local structure information of the graph network, which leads to the reduced inference ability of the node vectors in the space, and the final accuracy is lower than that of the other two methods, thus limiting the effectiveness of information dissemination. Therefore, the experiment results show that the use of higher-order information dissemination methods based on community structure can significantly improve the accuracy of the graph alignment task while guaranteeing the effectiveness, and at the same time significantly reduce the time required to learn the information about the global structure of the graph.

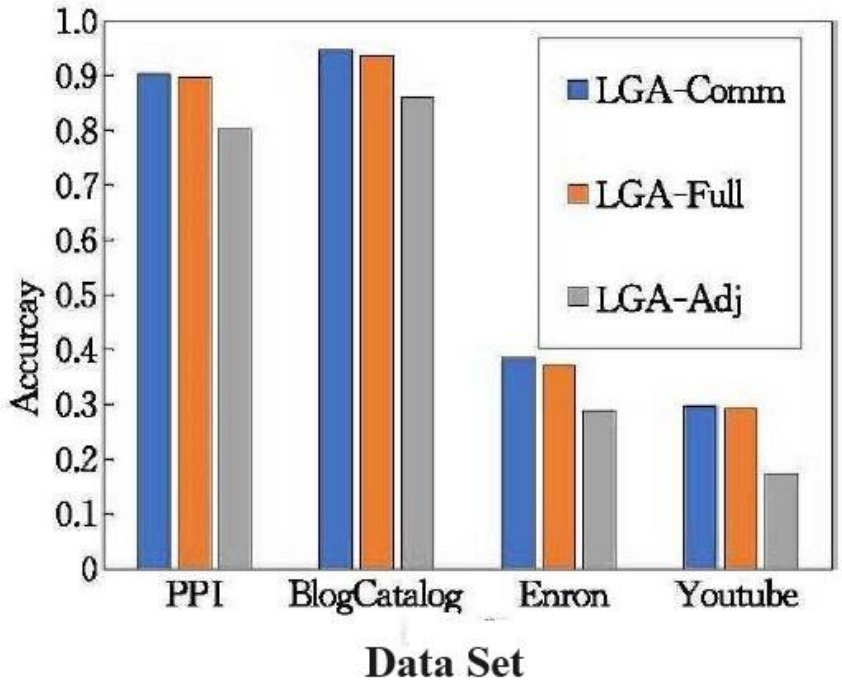

Figure 4: Results of ablation experiments

Table 3: Comparison of time consuming (seconds) for higher order information dissemination in different ways

NAME	LGA-COMM	LGA-ADJ
PPI	0.87 seconds	0.07 seconds
BlogCatalog	4.92 seconds	0.24 seconds
ENRON	5.18 seconds	0.41 seconds
YOUTUBE	59.48 seconds	4.28 seconds

1. PPI:

LGA-Comm: 0.87seconds

LGA-Adj: 0.07seconds

Analysis: In the PPI protein network dataset, the LGA-Adj approach took significantly less time to propagate higher-order information than the LGA-Comm approach. This suggests that for this particular dataset, the use of the LGA-Adj approach can propagate higher-order information more efficiently and less time-consumingly.

2. BlogCatalog:

LGA-Comm: 4.92seconds

LGA-Adj: 0.24seconds

Analysis: In the BlogCatalog social network dataset, it is also observed that the higher-order information dissemination time consumed by the LGA-Adj approach is much lower than that of the LGA-Comm approach. This implies that for social network data, the LGA-Adj approach may be a more efficient dissemination method.

3. Enron:

LGA-Comm: 5.18 seconds

LGA-Adj: 0.41 seconds

Analysis: In the Enron social network dataset, the higher-order message propagation time consumed by the LGA-Adj approach is still significantly lower than that of the LGA-Comm approach. This suggests that the LGA-Adj approach is also more efficient in this communication network data.

4. Youtube:

LGA-Comm: 59.48 seconds

LGA-Adj: 4.28 seconds

Analysis: In the Youtube News Network dataset, the LGA-Adj approach still takes significantly less time to propagate higher-order information than the LGA-Comm approach. Although compared with other datasets, both approaches take longer, the LGA-Adj approach still performs better in this large-scale network.

From these data, it can be seen that the LGA-Adj approach shows lower elapsed time for higher-order information dissemination on different datasets. This may mean that the LGA-Adj approach is more efficient in most cases, especially on larger network datasets. Researchers can choose the appropriate higher-order information dissemination approach based on the characteristics of their research questions and datasets to improve the efficiency and performance of the system. The results provide valuable information for the design and performance optimization of physical education teaching aid resource systems.

VI. Conclusion

In this paper, we propose an efficient graph representation learning method that combines sparse matrix decomposition and personalized random wandering, which can efficiently learn large-scale networks, and the representation vectors obtained can capture the global structural information of graphs, which is suitable for graph alignment tasks. A new label migration model is also designed to tackle the shortcomings of traditional graph annotation techniques, which can effectively improve the quality of the annotation results. A large-scale graph alignment algorithm LGA based on seed node fusion is proposed, which is different from the previous method of supervised graph representation learning with seed nodes, which performs unsupervised graph representation learning directly on the fused graph and maps the graph data to be matched into a unified vector space, thus avoiding the problem of low matching accuracy due to the transformation of the representation space.

Meanwhile, incorporating the sparse matrix decomposition technique can improve the generalization ability of the graph model, thus enhancing the accuracy of the large-scale graph alignment results. After extensive experimental validation, the method proposed in this paper shows excellent matching accuracy and computational efficiency on multiple types and different sizes of data, which is particularly suitable for efficiently aligning large-scale graph data. Utilizing the topological structure information of graphs, this paper proposes a comprehensive approach to achieve graph alignment, which has a wide range of application areas. However, the attribute information embedded in some graph data is extremely rich, which plays a crucial role in improving the accuracy of graph alignment. Therefore, in our future work, we will continue to improve the algorithm of this paper to better incorporate the network attribute information so as to further improve the accuracy of graph alignment.

With the continuous development of educational technology, the demand of educators and learners for diversified and personalized physical education teaching support resources is increasing. Designing and realizing a resource system with unsupervised representation learning function can meet this demand and bring more possibilities to the education field. This study explores in depth the potential of unsupervised representation learning technology in the field of education.

Through unsupervised representation learning, the sports system is able to automatically learn and extract useful information from sports teaching resources, so as to better satisfy users' needs and improve the intelligent recommendation and personalized adaptation of resources. We conducted a series of experiments to verify the effectiveness of the designed physical education teaching aid resource system in terms of resource recommendation, user satisfaction and learning effect. The experimental results show that the introduction of unsupervised representation learning significantly improves the performance of the system and receives positive feedback from users.

Although this study has achieved some results in the design of the physical education teaching aid resource system, there are still some challenges and future directions. Future research can further explore how to combine deep learning, natural language processing and big data technologies to further enhance the intelligence and performance of the system.

In summary, the findings of this study emphasize the importance and potential of unsupervised representation learning-based physical education teaching aid resource systems in the field of education to provide better teaching and learning experiences for educators and learners. We expect that this study will stimulate more research and

innovation on teaching-assisted resource systems for physical education and promote continuous progress and development in the field of education.

REFERENCES

[1] M. B. Zhang, Qi Qin, "Research on the Construction Method of Intelligence System Based on Division of Work and Collaboration in the Big Data Environment," *Journal of Intelligence*, vol. 41, no. 2, pp. 29-34, 2022.

[2] L. H. Zhu, "Research on Image Processing Optimization Technology of Deep Learning Algorithms Based on Big Data Processing Technology - Review of 'Introduction and Improvement of Computer Image Processing'," *Modern Radar*, vol. 43, no. 1, pp. 94, 2021.

[3] T. M. Cui, W. Liu, "Design of personalized online education system based on big data technology," *Modern Electronic Technology*, vol. 44, no. 5, pp. 175-180, 2021.

[4] J. Luo, "Design of Smart City Public Resource Allocation System Based on Big Data Technology," *Modern Electronic Technology*, vol. 44, no. 2, pp. 122-126, 2021.

[5] J. Z. Li, "Research on Big Data Audit Method Based on Data Warehouse Technology," *The Chinese Certified Public Accountant*, no. 2, pp. 66-68+3, 2022.

[6] H. P. Zhai, X. K. Shang, L. L. Han, "Research on Network Security Analysis Based on Big Data Technology," *Modern Electronic Technology*, vol. 45, no. 16, pp. 93-98, 2022.

[7] K. Zhou, Y. Q. Xiong, "Research on Urban Consumption Scenario Recognition Based on Unsupervised Clustering Method: Taking Changsha as an Example," *Modern Urban Research*, vol. 37, no. 10, pp. 32-39, 2022.

[8] H. Ding, W. Q. Ren, G. H. Cao, "Research on Academic Literature Representation Learning Based on Unsupervised Graph Neural Network," *Journal of the China Society for Scientific and Technical Information*, vol. 44, no. 1, pp. 11, 2022.

[9] Z. Z. Sun, H. H. Song, J. Q. Fan, Q. S. Liu, "Unsupervised Weak Light Image Enhancement Based on Global Local Generative Adversarial Learning," *Journal of Computer-Aided Design & Computer Graphics*, vol. 34, no. 10, pp. 1550-1558, 2022.

[10] Y. Liu, Z. Ma, Z. Y. You, et al. "Unsupervised clustering analysis of candidate signals for large-scale pulsars," *Acta Astronomica Sinica*, vol. 63, no. 3, pp. 129-138, 2022.

[11] C. M. Zou, D. Chen, "Unsupervised anomaly detection method for high-dimensional big data analysis," *Computer Science*, vol. 48, no. 2, pp. 121-127, 2021.

[12] X. B. Tang, M. R. Dong, R. Xu, "Design and Application of Text Mining Technology for CPI Prediction in the Context of Big Data," *Statistical Research*, vol. 38, no. 8, pp. 146-160, 2021.

[13] G. H. Cui, H. Shan, "Design of a Network Legal Information Resource Sharing System Based on Big Data," *Modern Electronic Technology*, vol. 44, no. 20, pp. 7-11.

[14] K. N. Huang, Y. Q. Guo, J. Yang, "Design of Optical Flow Field Image Matching Method for Big Data Mining Technology," *Laser Journal*, vol. 42, no. 5, pp. 107-111, 2021.

[15] M. Zhang, Y. C. Zhang, "Research on Building a Teaching Quality Monitoring System for Vocational Colleges Based on Big Data Technology," *Chinese Vocational and Technical Education*, no. 35, pp. 19-23, 2021.

[16] Y. H. Liu, "Intelligent Teaching System Based on Big Data Analysis Technology," *Modern Electronic Technology*, vol. 44, no. 7, pp. 178-182, 2021.

[17] M. Chang, "Design of Static Software Defect Detection System Based on Big Data Technology," *Modern Electronic Technology*, vol. 44, no. 17, pp. 37-41, 2021.

[18] M. R. Chen, "The Impact of Media Integration Technology on College English Education and Teaching: A Review of 'Teaching Reflection Methods and Techniques Based on Big Data of Classroom Teaching Behavior'," *China Sciencepaper*, vol. 16, no. 7, pp. 820, 2021.

[19] X. D. Zhou, L. Mei, "Design of an Intelligent Evaluation System for Mental Health Based on Big Data Technology," *Modern Electronic Technology*, vol. 44, no. 14, pp. 95-99, 2021.

[20] F. Tang, G. Y. Lan, "Big Data Governance Technology and Optimization of the Relationship between Central and Local Government: An Analysis Based on the Perspective of Information Politics," *Journal of Public Administration*, vol. 16, no. 2, pp. 138-156+199, 2023.

[21] C. W. Han, Z. D. He, et al. "Pairs trading via unsupervised learning - ScienceDirect," *European Journal of Operational Research*, 2022.

[22] Li L, Li X, Yang S, et al. "Unsupervised-Learning-Based Continuous Depth and Motion Estimation With Monocular Endoscopy for Virtual Reality Minimally Invasive Surgery," *IEEE transactions on industrial informatics*, vol. 6, pp. 3920-3928, 2021.

[23] Yusuke S, Yousuke I, Piljong J, et al. "Unsupervised learning architecture for classifying the transient noise of interferometric gravitational-wave detectors," *Scientific reports*, vol. 12, no. 1, 2022.

[24] Strigo I A, Simmons A N, Giebler J, et al. "Unsupervised learning for prognostic validity in patients with chronic pain in transdisciplinary pain care," *Scientific reports*, vol. 13, no. 1, 2023.

[25] G. JonnTerje. "Editorial for 'Transfer Learning Strategy Based on Unsupervised Learning and Ensemble Learning for Breast Cancer Molecular Subtype Prediction Using Dynamic Contrast Enhanced MRI'," *Journal of Magnetic Resonance Imaging*, vol. 55, no. 5, 2021.

[26] G. M. Wang, Chi Zhang, H. S. Wang, et al. “Unsupervised Learning of Depth, Optical Flow and Pose With Occlusion From 3D Geometry,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 99, 2020.
[27] V. Thambawita, B. T. Haugen, H. M. Stensen, et al. “P–029 Identification of spermatozoa by unsupervised learning from video data,” *Human Reproduction*, vol. 36, no. S1, 2021.
[28] Qi Ying; Zhou Shangbo; Zhang Zihan; Luo Shuyue; Lin Xiaoran; Wang Liping; Qiang Baohua, “Deep unsupervised learning based on color un-referenced loss functions for multi-exposure image fusion,” *Information Fusion*, vol. 66, no. 1, 2021.
[29] Ziolek RM, Smith P, Pink D L, et al. “Unsupervised Learning Unravels the Structure of Four-Arm and Linear Block Copolymer Micelles,” *Macromolecules*, vol. 54, no. 8, 2021.
[30] Nishi Y, “Non-Bernoulli operations of stochastic binary synapses using resistive switching devices for unsupervised learning in a spiking neural network,” *Applied physics express*, vol. 15, no. 7, 2022.