

Daradkeh, Mohammad

**Article**

## A user segmentation method in heterogeneous open innovation communities based on multilayer information fusion and attention mechanisms

Journal of Open Innovation: Technology, Market, and Complexity

**Provided in Cooperation with:**

Society of Open Innovation: Technology, Market, and Complexity (SOItmC)

*Suggested Citation:* Daradkeh, Mohammad (2022) : A user segmentation method in heterogeneous open innovation communities based on multilayer information fusion and attention mechanisms, Journal of Open Innovation: Technology, Market, and Complexity, ISSN 2199-8531, MDPI, Basel, Vol. 8, Iss. 4, pp. 1-19, <https://doi.org/10.3390/joitmc8040186>

This Version is available at:

<https://hdl.handle.net/10419/274487>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>



Article

# A User Segmentation Method in Heterogeneous Open Innovation Communities Based on Multilayer Information Fusion and Attention Mechanisms

Mohammad Daradkeh <sup>1,2</sup>

<sup>1</sup> College of Engineering and Information Technology, University of Dubai, Dubai 14143, United Arab Emirates; mdaradkeh@ud.ac.ae

<sup>2</sup> Faculty of Information Technology and Computer Science, Yarmouk University, Irbid 21163, Jordan; mdaradkeh@yu.edu.jo

**Abstract:** The heterogeneity and diversity of users and external knowledge resources is a hallmark of open innovation communities (OICs). Although user segmentation in heterogeneous OICs is a prominent and recurring issue, it has received limited attention in open innovation research and practice. Most existing user segmentation methods ignore the heterogeneity and embedded relationships that link users to communities through various items, resulting in limited accuracy of user segmentation. In this study, we propose a user segmentation method in heterogeneous OICs based on multilayer information fusion and attention mechanisms. Our method stratifies the OIC and creates user node embeddings based on different relationship types. Node embeddings from different layers are then merged to form a global representation of user fusion embeddings based on a semantic attention mechanism. The embedding learning of nodes is optimized using a multi-objective optimized node representation based on the Deep Graph Infomax (DGI) algorithm. Finally, the k-means algorithm is used to form clusters of users and partition them into distinct segments based on shared features. Experiments conducted on datasets collected from four OICs of business intelligence and analytics software show that our method outperforms multiple baseline methods based on unsupervised and supervised graph embeddings. This study provides methodological guidance for user segmentation based on structured community data and semantic social relations and provides insights for its practice in heterogeneous OICs.

**Keywords:** user segmentation; open innovation communities (OICs); heterogeneous networks; attention mechanisms; representation learning



**Citation:** Daradkeh, M. A User Segmentation Method in Heterogeneous Open Innovation Communities Based on Multilayer Information Fusion and Attention Mechanisms. *J. Open Innov. Technol. Mark. Complex.* **2022**, *8*, 186. <https://doi.org/10.3390/joitmc8040186>

Received: 14 September 2022

Accepted: 12 October 2022

Published: 14 October 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In today's fast-paced and ever-changing business environment, the traditional closed innovation paradigm is reaching its limits due to rapid technological advances, shortening product life cycles, and divergent consumer needs and preferences [1]. To advance their innovation-driven growth strategies, companies are gradually shifting from relying solely on R&D capabilities and internal resources to leveraging a variety of knowledge and external resources by creating an open innovation paradigm. Capitalizing on internal and external intelligence and communication channels, Chesbrough [2] describes open innovation as "the use of purposeful knowledge inflows and outflows to accelerate internal innovation and expand the use of innovation in external markets, respectively". This paradigm shift in innovation praxis, coupled with the widespread adoption of Web 2.0, has led to the development of open innovation communities (OICs), which are increasingly becoming a key driver for many organizations to foster their open innovation capabilities and to crowdsource new ideas and innovative solutions [3–5]. Recognizing their valuable contribution to innovation development and performance, a growing number of companies, such as Microsoft, Tableau, and Google, have created their own OICs focused on stimulating

users to contribute ideas and innovative solutions to their communities [3]. The key strength of OICs lies in their capacity to connect and incorporate heterogeneous knowledge bases where multiple and diverse groups of external stakeholders collaborate to explore, develop, and test new products, services, processes, or business models [6,7].

Because OICs are typically characterized by the heterogeneity of users and external knowledge resources, the business dilemma of segmenting users in heterogeneous OICs remains one of their inherent and most controversial issues [5,8–10]. Typically, the OIC is a network structure that consists of a stack of various interactions between community users, including several layers of community-structured data and rich semantic and social relationships. Users engage with each other, with subjects, and with ideas in an OIC, forming a close-knit network of relationships that is often defined in terms of several degrees of community structure. A heterogeneous OIC includes two types of nodes, users and ideas, and various relationships between them, including users providing ideas, users commenting on ideas, and users voting on or against ideas. Extracting user node embeddings from heterogeneous OICs and segmenting the user base are key aspects of providing disciplinary services such as personalized recommendations and user analytics predictions [11]. A robust and highly accurate predictive model is imperative to improve recommendation prediction and assist community operators in making informed decisions, especially when they are overwhelmed by large and dispersed heterogeneous user communities and their interactions with the correspondingly large and diverse knowledge base. Therefore, analyzing user communities and developing user segmentation methods in heterogeneous OICs is becoming increasingly important in order to extract the best value from shared content, improve innovation performance, and increase the business return on R&D resources [7,12,13].

The extant research on open innovation and platforms has developed several methods for user segmentation; however, most of these methods either use only one social relationship or are based on explicit community relationships that are not always present [14–16]. Recently, several studies have emphasized the importance of incorporating implicit relationships to improve the accuracy of user segmentation results [17–19]. However, these studies rely on the basic premise that community relationships are homogeneous; that is, the nodes and relationships in these communities are of the same type. Typically, in OICs, different relationships exist between various types of participants and stakeholders [18]. For example, there are relationships of trust, interest, and friendship that exist only in one direction. Similarly, users follow each other because they have the same interest in an idea or topic, but they may have different expertise and knowledge about a topic. Therefore, using an approach based on equality and similar affiliation of user nodes may produce inaccurate segmentation results. To address these issues, several recent studies have developed user segmentation models to capture the diversity and heterogeneity of OICs [18,20,21]. Multiple entities and edge types in heterogeneous OICs provide a large amount of information that can be effectively used to mitigate sparsity effects and improve decision efficiency. In addition, information in heterogeneous OICs can be used to capture implicit relationships among community members [22,23]. However, the identification and incorporation of implicit relationships in the user segmentation process remain largely unexplored in the open innovation literature.

With the development of various graph technologies in recent years, researchers have shown increasing interest in segmenting user communities using node embedding-based methods. As a result, various machine learning techniques have been effectively applied for user segmentation in OICs [24–26]. Based on different user node embedding and graph representation learning methods, the current approaches can be broadly classified into random walk-based methods, such as DeepWalk [27] and Node2Vec [28], and graph convolutional network (GCNs)-based methods [26,29]. The random walk-based user segmentation method can support large-scale scenarios; however, due to the stochastic nature of the walking strategy, the results generated by user segmentation will not be optimized. Although Node2Vec can optimize the random walk strategy, the user segmentation based

on the Node2Vec algorithm ignores the edge type information of the heterogeneous network in which the user nodes are located, resulting in poor accuracy of the community segmentation results. On the other hand, graph convolutional network (GCN)-based user segmentation methods can handle both user node features and community structure features; however, these methods are difficult to extend to large-scale heterogeneous community scenarios, such as the case of OICs [18,30].

In view of the shortcomings of previous work, this study proposes a new user segmentation method based on multilayer information fusion and semantic attention mechanisms, aiming to improve the accuracy of segmentation prediction and alleviate the data sparsity problem in OICs. The method stratifies heterogeneous OICs based on different edge types and combines semantic information among the layers to improve the accuracy of user segmentation results. After representing node embeddings in a single-layer network, the semantic information between the layers is merged to obtain a fused representation of user nodes. Finally, the k-means clustering algorithm is used for user segmentation based on user embeddings. To test our proposed approach, we collected several datasets from four OICs, namely, the Microsoft Power BI community, Tableau community, Qlik community, and RapidMiner community. These datasets were collected from August 2021 to August 2022. The contributions of this study are summarized as follows:

- This study investigates the user segmentation problem in heterogeneous OICs and develops a hierarchical processing method to transform heterogeneous communities into multiple heterogeneous networks in an attempt to better distinguish and fuse network structure information and semantic information and improve the accuracy of community segmentation.
- This study extends the optimization function of the multi-objective Deep Graph Infomax (DGI) [31] algorithm to control the similarity of the community structures explored from different data sources; therefore, the effect of noise can be reduced. In addition, we combine the structural features of heterogeneous OICs with the semantic features of user nodes to accurately construct user node embeddings in a single-layer network.
- This study compares our method with multiple baseline methods based on unsupervised and supervised graph embedding techniques using a real-world dataset collected from OICs developed for business intelligence and analytics tools and stakeholders. Further ablation experiments were conducted to evaluate the effectiveness of different parts of the proposed method.

The rest of the paper is organized as follows: Section 2 provides a review of related studies. Section 3 provides a detailed description of the proposed method. Section 4 presents the results of the experimental analysis and performance evaluation. Section 5 summarizes and discusses the main findings of this study. Finally, Section 6 presents the limitations of this study and highlights the directions for future work.

## 2. Related Works

### 2.1. Open Innovation Communities

The concept of open innovation originates from Chesbrough's proposition on open innovation [2], which states that firms should change the way innovation is developed and nurtured, evolving from more traditional idea generation mechanisms to new forms of open innovation that involve both internal and external stakeholders in idea generation. From a practical perspective, open innovation is described as an Internet-based collaborative innovation platform that leverages the innovative contributions of external actors. External actors are further described as heterogeneous groups of non-commercial actor constellations whose members are informally and voluntarily involved in the collaborative generation, development, and application of new knowledge and innovative products [23,32]. In this sense, the heterogeneity of users and the diversity of external knowledge resources make the open innovation process a gradual and cumulative engagement process, where accumulation drives the division of user groups and at the same time

defines their boundaries. According to Chesbrough et al. [33], community-based open innovation is usually based on the conceptualization of utilitarian interrelationships of interest, i.e., in the social rather than the communicative domain. Therefore, the processing of information and knowledge are stylistic determinants of the coordinated innovation process, as they are the core resources for the development of new ideas and innovative products. Due to its informal nature, the open community-based model of coordination differs significantly from the dominant hierarchical or market-like approach to exchange relations in the economic context, where conventions are usually used to define who must do what and how to deal with the knowledge generated [32]. In this context, openness and heterogeneity constitute necessary prerequisites for the reproduction of community relations and are recognized by all actors as shared values and guiding directions for their contribution to community activities [1,34,35].

In essence, OICs encompass multi-heterogeneous networks whose growing importance stems from the general digitization of social interactions and communication processes. By browsing, commenting, or voting on other users' ideas, users form relationships with other users and recognize other users who share their interests or appeal to them as objects of interest, thus devoting more attention to their contributions and participation [33]. Moreover, users' ideas are expressed in the community mainly through text and images. The content of each idea consists of one or more keywords, and different contents attract users with the same interests and preferences. Different users' ideas are connected through the same content, thus forming a knowledge network based on the core content. The social network of users and the content knowledge network are closely connected through the innovative activities of the community. Users form ideas, and ideas attract users; thus, the innovation activities of the open innovation community ecosystem are constantly iterated and developed [34,36].

In heterogeneous OICs, user-multiple social networks and idea-multiple networks are connected by different types of bidirectional interaction edges, such as user-idea viewing, user-idea contribution, and user-idea evaluation. Multiple heterogeneous networks are a combination of heterogeneous networks [30] and multiple networks, where multiple networks are represented as shown in Figure 1a, which consists of multiple network layers, each of which is a mere network with the same set of nodes. The node set contains multiple nodes of the same type, but each layer of the simple network includes only one type of edge, and the types of edges differ between the layers. As shown in Figure 1b, the heterogeneous network itself is a multilayer network, where each layer is a simple network with specific types of nodes and interacting edges, and two simple networks are connected by bidirectional interacting edges, i.e., different types of nodes belonging to two simple networks are connected by interacting edges. Figure 1c depicts a multi-heterogeneous network consisting of two multiple networks [33], where the multiple networks are connected by different types of bidirectional interaction edges, specifically described as a layer of one of the multiple networks associated with a layer of the other multiple network.

## 2.2. User Segmentation in OICs

With the rapid development of open innovation models and the increasing complexity of heterogeneous networks, the study of user segmentation in open innovation has received a great deal of attention from academia and industry. Broadly classified, user segmentation methods are mainly divided into module optimization algorithms [18], tensor factorization algorithms [37], label propagation algorithms [38], and node embedding algorithms [39]. Module optimization-based algorithms extend modularity maximization from single-layer networks to multi-layer networks; however, these algorithms inherently ignore the fact that different relations have different importance. The tensor decomposition-based algorithms use a search heuristic framework to obtain the optimal number of clusters; however, the performance of these algorithms degrades in the presence of a large number of user groups. Label propagation-based algorithms use label information for user segmentation;

however, these algorithms may produce community labels that do not match the real node attributes [40]. Compared with previous types of user segmentation algorithms, user node embedding-based algorithms can better preserve the complex information contained in the network and provide better accuracy of user segmentation results [18,41].

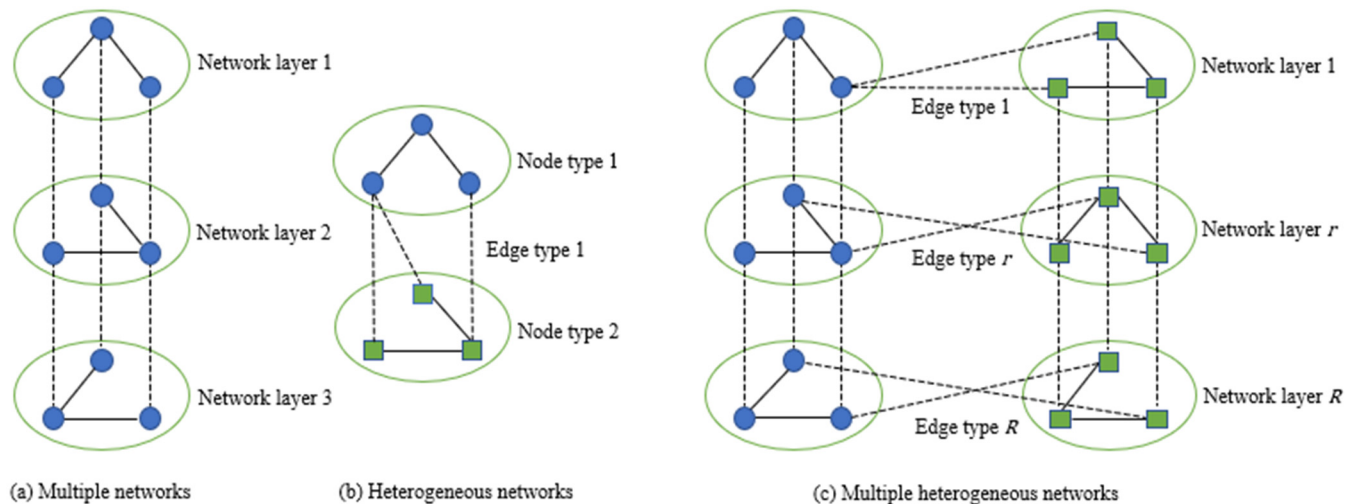


Figure 1. Different types of networks.

Most existing research on user segmentation methods relies on explicit social relationships, which are not always present in OICs, and even when explicit relationships exist, the data are typically sparse and noisy [30,42]. In response, researchers have recently emphasized the inclusion of implicit social relationships in user segmentation methods. For example, Su et al. [43] proposed a link prediction model based on Dempster–Shafer theory to compute implicit relationships of users in social recommender systems. Daneshvar and Ravanmehr [44] developed an idea recommendation algorithm model based on implicit trust relationship inference while incorporating temporal features into the system. Liu and He [45] used trust propagation and aggregation strategies to identify indirect trust of users in OICs. Ahmadian et al. [46] used link prediction techniques to extract implicit relationships in OICs and proposed an explicit- and implicit-based friendship links approach. The aforementioned scholars inferred implicit social relationships from explicit social relationships, while others identified implicit social relationships through user-item rating metrics [7,12]. Huang et al. [47] proposed that the top  $k$  similar users of each user can be identified by calculating the Pearson correlation coefficient between each user. Awati and Shirgave [48] suggested using the Hellinger distance to extract the user-implied relationships in the idea bipartite graph.

Despite the valuable contributions of existing user segmentation methods in OICs, their performance is still considered inadequate when dealing with complex networks containing heterogeneous community structures. In heterogeneous OICs, users are usually connected to each other through different types of nodes and link relationships; in this case, multiple types of entities and edges can be utilized to improve the accuracy of user segmentation. Schutera et al. [49] developed a fuzzy overlapping community segmentation algorithm based on node vector representation using graph convolution to achieve efficient community segmentation in complex networks. Jia et al. [50] modeled different types of interactions between nodes and links and transformed the idea recommendation problem into a node proximity computation problem on a heterogeneous graph. Xu et al. [51] modeled the entire user community as a heterogeneous information network and computed user similarity by learning from meta-path-based embedding representations to identify hidden friends. Han et al. [52] proposed a social network link community segmentation algorithm based on  $k$ -means and user node embedding and multi-layer information fusion analysis to accurately identify overlapping communities in a recommender system.

Wang et al. [53] suggested using the random walk method to establish node adjacencies on a heterogeneous graph and then model the structural and semantic relationships of the heterogeneous network. Weng et al. [30] constructed a collaborative filtering recommendation algorithm based on multilayer information fusion and relational clustering and applied it to collaborative filtering of user segmentation to solve the problem of high sparsity and high dimensionality of data in OICs.

The previous work discussed above shows that multilayer information fusion, node embedding analysis, and attention mechanisms have great potential for user segmentation in OICs. Therefore, in this study, user node embedding analysis was used to compute similarities among users in heterogeneous OICs and thus recommend ideas that are closely related to the interests and priorities of the target users. Moreover, both the proposed method and the baseline methods used in the comparative evaluation performance experiments belong to user segmentation algorithms based on user node embedding.

### 3. Proposed Method

The user segmentation method proposed in this study uses a hierarchical processing approach to transform the heterogeneous OIC into a multi-heterogeneous network, aiming to better distinguish and fuse network structure and semantic information to improve the accuracy of community segmentation. The multi-heterogeneous network of OIC includes two types of nodes, users and ideas, and various relationships between them, such as users' browsing ideas, users' contributing ideas, and users' commenting and liking ideas. The proposed method stratifies the heterogeneous network based on different relationship types and obtains the representation of user nodes from a single layer of the network. Then, the user embeddings of the same user node in different layers are fused based on the semantic attention mechanism to obtain the user fusion embedding representation. In addition, the global optimization of the user fusion representation is achieved using a multi-objective DGI (Deep Graph Infomax) algorithm [45] by maximizing a mutual information objective function consisting of the combination of the global summary vector  $s$ , the feature attributes  $f_i$  of the user nodes, and the user fusion embedding representation. Finally, the k-means algorithm is used to form user clusters based on the influence of user nodes. The basic framework of the proposed user segmentation method is shown in Figure 2.

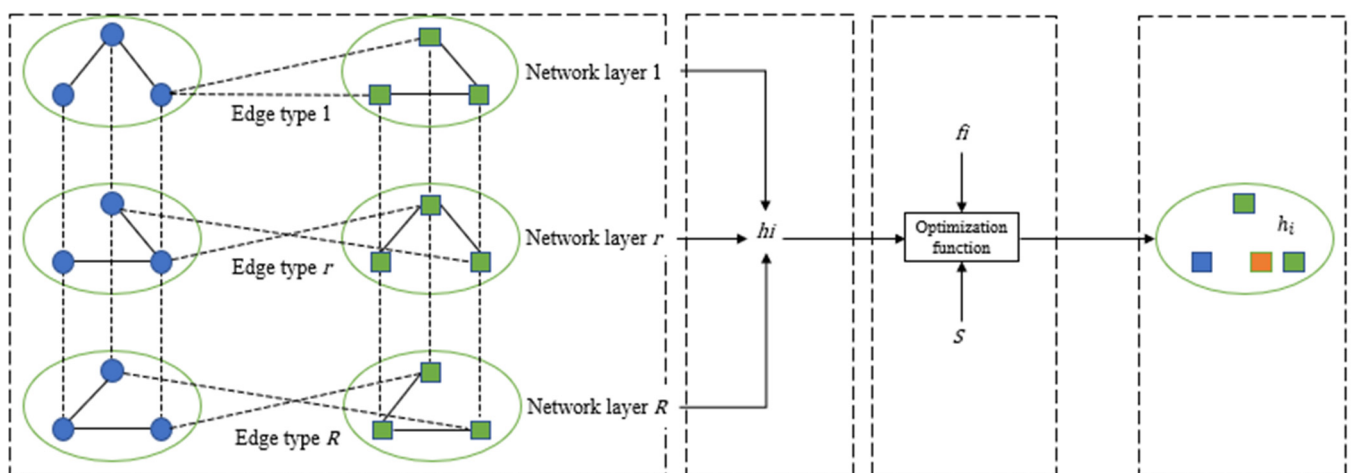


Figure 2. Framework of user segmentation in OICs.

#### 3.1. User Node Embedding

The heterogeneity of OIC encompasses multiple types of users–ideas interactions. In this study, the heterogeneous OIC is layered according to different interaction types. First, the features of user nodes are embedded in a single layer of the heterogeneous network under a certain interaction type. This is achieved by combining the structural features of OIC heterogeneity and the semantic features of user nodes to accurately construct the embed-

ding of user nodes in the single-layer network. In this study, the GATNE-T algorithm [54] was used to define the overall embedding of user nodes based on edge types and lay the foundation for subsequent feature analysis and fusion of multiple heterogeneous networks. Specifically, the overall embedding of user node  $v_i$  on each edge type  $r$  is divided into two parts: the basic embedding, and the edge embedding. The basic embedding of node  $v_i$  is shared among different edge types. The algorithm aggregates the edge embeddings of node  $v_i$  based on edge type  $r$  after aggregating neighboring node embeddings with  $k$  iterations  $u_{i,r}^{(k)} \in R^S (1 \leq k \leq K)$ , as shown in Equation (1).

$$u_{i,r}^{(k)} = aggregator \left( u_{j,r}^{(k-1)}, \forall v_j \in N_{i,r} \right) \tag{1}$$

The edge embedding  $u_{i,r}^{(k)}$  after the  $k_{th}$  iteration is used as the final representation of the edge embedding  $u_{i,r}$ , and the different edge embeddings in the network corresponding to a particular node  $v_i$  are combined into a matrix  $U_i$ , as shown in Equation (2).

$$U_i = (u_{i,1}, u_{i,2}, \dots, u_{i,m}) \tag{2}$$

The overall embedding model representation of the user node  $v_i$  on edge type  $r$  obtained by this algorithm is shown in Equation (3).

$$v_{i,r} = h_z(f_i) + \alpha_r M_r^T U_i a_{i,r} \tag{3}$$

where  $h_z(f_i)$  denotes the basic embedding of user node  $v_i$ .

**Definition 1.** User node embedding is defined as the transformation function of the feature attribute  $f_i$  of user node  $v_i$ , and  $a_{i,r} \in R^m$  are the significance coefficients of different types of edge embeddings. In our method, network structural features and attribute features are projected into the heterogeneous information network, and then the generated network vertices are embedded using the GATNE-T algorithm [54], which is an inductive embedding algorithm consisting of two components: edge embedding, and node attributes.

The above processing is used as an initial representation of user node embedding on a single-layer heterogeneous network in the OIC.

### 3.2. Representation Fusion

In the multi-heterogeneous OIC, the network based on a bidirectional interaction edge type  $r \in R$  is referred to as the  $r_{th}$  layer of the network. In the  $r_{th}$  layer network, the embedding of user node  $v_i$  can be further represented as  $h_i^r$ . This study used the overall embedding model of user nodes proposed in the previous section as the network encoder  $\epsilon$ . The specific description is given as a network of the  $r_{th}$  layer, where the embedding of the user node  $v_i$  is represented as shown in Equation (4).

$$h_i^r = v_{i,r} = h_z(f_i) + \alpha_r M_r^T U_i a_{i,r} \tag{4}$$

Considering the relevance of different network layers in multiple heterogeneous networks, this study used a semantic attention mechanism [50,55] to fuse the embedding of user nodes at different layers, as shown in Figure 3.

For the  $r_{th}$  layer of the network, the layer weight  $\alpha_i^r$  of user node  $v_i$  is obtained using a layer-based semantic attention mechanism, as shown in Equation (5).

$$\alpha_i^r = \tanh \left( (y^r)^T V_r h_i^r \right) \tag{5}$$

where  $V_r \in R^{d' \times d}$  is the parameter matrix,  $y^r$  denotes the hidden representation vector of the  $r_{th}$  layer of the network, and  $\tanh$  denotes the tangent activation function.



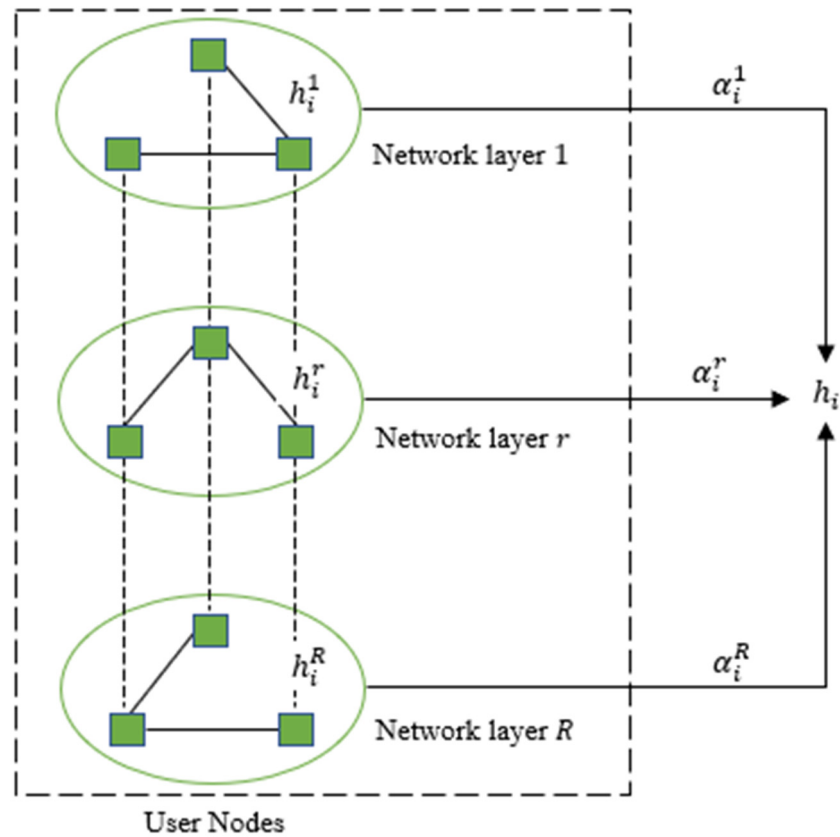


Figure 3. Representation fusion.

The node embedding weights of different layers are normalized, and the results are obtained as shown in Equation (6).

$$\alpha_i^r = \frac{\exp(\alpha_i^r)}{\sum_{r'=1}^R \exp(\alpha_i^{r'})} \tag{6}$$

where  $R$  denotes the number of layers of the OIC multi-heterogeneous network.

The resulting fused embedding representation of user nodes  $v_i$  in the multiple heterogeneous networks of OIC is shown in Equation (7).

$$h_i = \sum_{r=1}^R \alpha_i^r h_i^r \tag{7}$$

### 3.3. Parameter Optimization

The fused embedding representations of user nodes obtained in the previous section are used as initial inputs to the optimization part of the DGI [31] algorithm, considering only the external supervision signals, i.e., the mutual information between the fused embedding representation  $h_i$  of user node  $v_i$  and the global summary vector  $s$  in the network. The optimization part of the algorithm contains only information about the association of a particular user node with other user nodes. The internal supervision signal, i.e., the fusion embedding of user node  $v_i$  representing the mutual information between  $h_i$  and the feature attribute  $f_i$  of that node, is not fully exploited and has the disadvantage of ignoring the feature attribute of the node itself. Therefore, in our method, the fused embedding representation of a user node is globally optimized by introducing higher-order mutual information [31] to capture the external and internal supervision signals and the synergy between them.

For the three parameters of the user node  $v_i$  in the OIC multi-heterogeneous network, i.e., the fused embedding representation of the user node  $h_i$ , the global summary vector  $s$ , and the node feature attribute  $f_i$ , the fused embedding representation of the user node is globally optimized by maximizing a mutual information objective function consisting of a combination of the three. This preserves the association between the fused embedding of a particular user node and the fused embeddings of other user nodes, as well as the feature attributes of the user nodes themselves. According to the definition of higher-order mutual information [56], when the number of random variables is 3, the mutual information between the distribution of  $X1$  and the joint distribution of  $X2$  and  $X3$  can be obtained as shown in Equation (8).

$$I(X1; X2; X3) = H(X1) + H(X2) - H(X1, X2) + H(X1) + H(X3) - H(X1, X3) - H(X1) - H(X2, X3) + H(X1, X2, X3) = I(X1; X2) + I(X1; X3) - I(X1; X2, X3) \tag{8}$$

Following the DGI process, in our method, the three random variables  $X1$ ,  $X2$ , and  $X3$  in Equation (8) were replaced with the three parameters  $h_i$ ,  $s$ , and  $f_i$  of the user node  $v_i$ , resulting in Equation (9).

$$I(h_i; s; f_i) = I(h_i; s) + I(h_i; f_i) - I(h_i; s; f_i) \tag{9}$$

where  $I(h_i; s)$  captures the internal supervision signal, i.e., the external mutual information between the fused embedding representation  $h_i$  of user node  $v_i$  and the global summary vector  $s$ .  $I(h_i; f_i)$  captures the internal supervision signal, i.e., the intrinsic mutual information between the fused embedding representation  $h_i$  of the user node  $v_i$  and the feature attribute  $f_i$  of that node.  $I(h_i; s; f_i)$  captures the interactions between the extrinsic and intrinsic supervision signals. In this study, we extended the optimization approach of the DGI algorithm and proposed to maximize  $I(h_i; s; f_i)$  for parameter optimization, i.e., we proposed to use the joint maximization of  $I(h_i; s)$ ,  $I(h_i; f_i)$ , and  $I(h_i; s; f_i)$  to obtain the optimization results, and the final fused embedding representation  $h_i$  of user node  $v_i$  was obtained through the joint optimization.

### 3.4. User Clustering

Based on the fused representation of user nodes in multiple heterogeneous networks in the OIC, this study proposed a computational method for user clustering. Given that the k-means algorithm [57] allows for relatively efficient and intuitive processing of large-scale data with high scalability, this study used the k-means algorithm for user clustering by incorporating node influence into centroid identification and adjusting the selection of initial clustering centroids as follows:

- The influence of user nodes is determined using the user node fusion representation.
- The obtained influence of user nodes is ranked in descending order, and the  $k$  user nodes with the highest influence are selected as the initial clustering centroids of the k-means algorithm.
- The k-means algorithm is iteratively applied until a stable user segmentation emerges.

Before deriving the influence of user nodes, the distance calculation of user nodes needs to be considered. Combining the standardized Euclidean distance can better balance the characteristics of independence and correlation between multiple dimensions. Therefore, in our method, the standardized Euclidean distance [18] was used to calculate the distance between user nodes  $v_i$  and  $v_j$  using the optimized fusion representation of user nodes as input. The influence of user nodes was calculated as shown in Equation (10).

$$P_{v_i} = \frac{1}{\sum_{v_j \in V} \sqrt{\sum_{k=1}^n \left(\frac{h_{ik} - h_{jk}}{S_k}\right)^2}} \tag{10}$$

where  $P_{v_i}$  denotes the influence of user node  $v_i$  in the multi-heterogeneous networks of OIC.

### 3.5. Algorithm Description

The algorithm proposed in this study extracts user node embeddings from each layer of the OIC multi-heterogeneous network. Then the user node embeddings from each layer of the network are merged to obtain the fused embeddings of the user nodes. Finally, the fused node embeddings are optimized by an objective function, which are then used as inputs to improve the k-means clustering algorithm. The general framework of the proposed user segmentation algorithm in heterogeneous OIC based on multilayer information and attention mechanisms is shown in Algorithm 1.

---

**Algorithm 1.** User segmentation algorithm in heterogeneous OIC based on multilayer information and attention mechanisms.

---

Input: OIC multi-heterogeneous network  $G_{MH} = (V, E, F)$ , number of network layers  $|R| > 1$ , number of user communities  $k$   
Output: User segmentation result  $C = (C_1, C_2, \dots, C_k)$

- (1) For each multi-heterogeneous network in layer  $r \in R$  network
- (2)     For each user node
- (3)         Obtain the user node embedding representation of at layer  $r$  using Equation (4)
- (4)         Obtain the layer weights of the user nodes using the layer-based semantic attention mechanism in Equation (5)
- (5)         Normalize the layer weights of user nodes using Equation (6)
- (6)     End for
- (7) End for
- (8) For each user node
- (9)     Obtain the fused embedding representation of the user node using Equation (7)
- (10)    Optimize the fused embedding representation of the user node using the objective function using Equation (9)
- (11) End for
- (12) Calculate the influence of user nodes and select the top  $k$  user nodes as the initial user community centers using Equation (10)
- (13) Use k-means algorithm for user segmentation

---

## 4. Experimental Analyses

### 4.1. Datasets

The datasets for this study were collected from four OICs of business intelligence and analytics software. The four OICs used in the experimental analysis were the Microsoft Power BI community (<https://community.powerbi.com>, accessed on 1 September 2022), the Tableau community (<https://community.tableau.com/s/>, accessed on 1 September 2022), the Qlik community (<https://community.qlik.com/>, accessed on 1 September 2022), and the RapidMiner community (<https://community.rapidminer.com/>, accessed on 1 September 2022). These OICs were selected based on their popularity and publicly available data related to the activities of the OICs' members and host companies. These OICs were created as online crowdsourcing platforms specifically to connect companies and users and solicit suggestions for solving problems or generating ideas and inspiration for new projects to test business intelligence and analytics tools and practices. These business intelligence and analytics tools provide business users with interactive visualization and analysis through an intuitive interface to create their own dashboards and analytics applications [3,58]. The OICs consist of users and customers from different countries, cultures, backgrounds, and expertise with a variety of business intelligence and analytics solutions. In order to participate, users can create a profile and join the community for free by using the email address provided by the community. When they post an idea, they must provide a title for the idea and a description of its topic and select the category to which the idea belongs. In addition, by posting ideas, members can interact with other users by voting, scoring, and reviewing other people's ideas.

The dataset was collected, stored using a web crawler, and processed using the Python language and statistical analysis between August 2021 and August 2022. The four datasets used in this experimental analysis contained different types of relationships between users and ideas, such as user idea views, user idea contributions, user idea comments, and ratings [1,35]. This experiment also used information about users and the ideas they contributed, including user attributes, idea attributes, and ratings of idea implementation.

In our experiments, we chose the set of meta-paths as  $P = (UVI; UCI)$  to label the two types of relationships (users viewing ideas, and users contributing ideas). Since the purpose of this study was to understand and analyze the user segmentation in OICs, the classification of ideas and the demographic characteristics of idea contributors were not considered. The datasets used in this experiment included the number of nodes, edges, node types, and edge types, but each dataset contained different node attributes, as shown in Table 1.

**Table 1.** Dataset comparison.

Datasets	Node Type	No. of Nodes	Edge Type	Network Layer Corresponding to Edge Type	No. of Edges
Power BI	User	2460	User viewing ideas	UVI	84,853
	Ideas	33,660	User contributing ideas	UCI	64,843
Tableau	User	8556	User viewing ideas	UVI	49,439
	Ideas	81,633	User contributing ideas	UCI	22,751
Qlik	User	1129	User contributing ideas	UVI	69,108
	Ideas	29,034	User contributing ideas	UCI	33,853
RapidMiner	User	3908	User viewing ideas	UVI	59,482
	Ideas	30,502	User contributing ideas	UCI	32,761

#### 4.2. Evaluation Indicators

To evaluate our method and different baseline methods, two classical performance evaluation metrics were used in user node clustering and user node similarity search experiments, namely, the normalized mutual information (NMI) between two clusters and the Sim@5 value [18,59]. For the node clustering experiments, a self-supervised signal training model followed by a modified k-means algorithm was used in order to obtain the NMI values. NMI is a metric used to measure the accuracy of community segmentation when real labels are available on the network. Its definition is given by Equation (11).

$$NMI(A, B) = \frac{-2 \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} M_{ij} \log\left(\frac{M_{ij}N}{M_i M_j}\right)}{\sum_{i=1}^{n_1} M_i \log\left(\frac{M_i}{N}\right) + \sum_{j=1}^{n_2} M_j \log\left(\frac{M_j}{N}\right)} \tag{11}$$

For the node similarity search experiment, the cosine similarity between each pair of user nodes is calculated according to [59,60]. Then, for each user node, the top 5 most similar user nodes are selected, and the proportion of these user nodes belonging to the same user community (class) is calculated; the accuracy of user node embedding is determined based on a value called Sim@5.

#### 4.3. Baseline Methods

The proposed USOIC was compared with two types of baseline algorithms, namely, unsupervised and supervised user segmentation algorithms:

##### (A) Unsupervised algorithms

- DeepWalk [61]: This method uses the Random-Walk strategy to obtain the node sequence; then, the Skip-Gram algorithm is used to obtain the node representations; finally, the objective function is optimized according to the hierarchical Softmax.
- Node2Vec [62]. This method is a more general abstract representation of the DeepWalk algorithm, which mainly improves the former Random-Walk strategy to obtain neighborhood information and more complex node dependencies.
- MetaPath2Vec [28]: This meta path-based method for embedding heterogeneous networks aims to deal with the heterogeneity of nodes. The MetaPath2Vec

algorithm degenerates to the DeepWalk algorithm when there is only one node type in the network.

- CommDGI [63]: This method is an unsupervised learning algorithm based on mutual information for dealing with homogeneous networks.

(B) Supervised algorithms

- GCN [30]: This method is a semi-supervised algorithm applied to node classification in homogeneous networks, which uses a convolution operation to merge the feature representation of neighbors into the node feature representation.
- GAT [64]: In this method, the attention mechanism is applied to homogeneous networks that require a supervised setup, and the algorithm learns node embeddings based on the local structure of the nodes.
- HAN [18]: This method uses node-level attention and semantic-level attention to capture information about all meta-paths.

**5. Performance Analysis and Evaluation**

Based on the four datasets of OICs, this study compared the results of different algorithms in node clustering and node similarity search experiments, as shown in Table 2.

**Table 2.** Baseline comparison of similarity search and node clustering tasks.

Dataset	Power BI		Tableau		Qlik		RapidMiner	
	NMI	Sim@5	NMI	Sim@5	NMI	Sim@5	NMI	Sim@5
Indicators								
DeepWalk	0.082	0.725	0.116	0.491	0.347	0.627	0.312	0.702
Node2Vec	0.073	0.737	0.122	0.486	0.381	0.626	0.308	0.711
MetaPath2Vec	0.085	0.746	0.128	0.491	0.386	0.633	0.316	0.713
CommDGI	0.006	0.556	0.182	0.577	0.552	0.784	0.642	0.887
GCN	0.286	0.623	0.175	0.564	0.464	0.722	0.672	0.865
GAT	0.302	0.631	0.182	0.551	0.467	0.724	0.665	0.871
HAN	0.028	0.493	0.162	0.562	0.471	0.776	0.655	0.871
Our Method-Average Pooling	0.342	0.743	0.187	0.602	0.556	0.774	0.683	0.874
Our Method	0.345	0.754	0.195	0.606	0.564	0.788	0.692	0.899

In both node clustering and node similarity search experiments, our method showed improvements in NMI and Sim@5 metrics compared to other benchmark node-embedding user segmentation methods. Our method combines structural information, semantic information, and user node independence information of multiple heterogeneous networks in OIC to learn effective user node embeddings and to obtain better user segmentation results. As a result, it typically learns better and produces more accurate user node embeddings compared to unsupervised and supervised algorithms. In addition, it produces higher user community density with significantly fewer edge user nodes between communities, with better stability and scalability.

Among the unsupervised algorithms used for comparison, the DeepWalk-based user segmentation algorithm performed poorly in both types of experiments because the algorithm could not properly handle the heterogeneity of OICs. The MetaPath2Vec-based user segmentation algorithm cannot handle multiple semantic information at the same time, which makes the validity of the user node embeddings obtained by this algorithm insufficient. Among the supervised algorithms used for comparison, the GCN-based user segmentation algorithm and the GAT-based user segmentation algorithm are close to the corresponding metrics of our algorithm in terms of NMI metrics, but there are large differences in terms of Sim@5 metrics, which proves that our algorithm can handle edge type information and semantic information more reasonably and obtain a higher density of user communities.

The method proposed in this study used ablation experiments to further evaluate the performance of user node fusion representation and the effectiveness of user segmentation. To evaluate the performance of the proposed representation fusion, the proposed approach using representation fusion was compared with the proposed algorithm using average pooling instead of representation fusion. As shown in the last two rows of Table 2, the former showed marginal improvements in both metrics.

Combined with the parameter optimization in the proposed method (Section 3.3), the optimization part was divided into the external supervision signal  $I(h_i; s)$  combining global summary vectors (referred to as  $E$ ), the internal supervision signal  $I(h_i; f_i)$  combining node feature attributes (referred to as  $I$ ), and the joint supervision signal  $I(h_i; s; f_i)$  combining global summary vectors and node feature attributes (referred to as  $J$ ). For the optimization part, we conducted ablation experiments, i.e., comparison experiments, by combining the external supervision signal ( $E$ ), the internal supervision signal ( $I$ ), the joint supervision signal ( $J$ ), and the reconstruction error of the single-layer network (referred to as  $R$ ) to demonstrate the optimization effect of the parameters proposed in this study, as shown in Table 3.

**Table 3.** Ablation experiments of our method in two types of experiments.

Dataset		Power BI		
Network Layer		UVI	UCI	
Indicators	NMI	Sim@5	NMI	Sim@5
E	0.002	0.395	0.003	0.414
E + R	0.002	0.399	0.003	0.426
E + I	0.152	0.512	0.143	0.512
E + I + J	0.163	0.566	0.153	0.593
Dataset		Tableau		
Network Layer		UVI	UCI	
Indicators	NMI	Sim@5	NMI	Sim@5
E	0.547	0.801	0.087	0.493
E + R	0.551	0.804	0.077	0.491
E + I	0.512	0.802	0.144	0.524
E + I + J	0.592	0.806	0.142	0.528
Dataset		Qlik		
Network Layer		UVI	UCI	
Indicators	NMI	Sim@5	NMI	Sim@5
E	0.526	0.626	0.651	0.812
E + R	0.525	0.659	0.659	0.833
E + I	0.527	0.728	0.655	0.872
E + I + J	0.527	0.708	0.656	0.874
Dataset		RapidMiner		
Network Layer		UVI	UCI	
Indicators	NMI	Sim@5	NMI	Sim@5
E	0.403	0.730	0.053	0.543
E + R	0.422	0.711	0.052	0.558
E + I	0.403	0.711	0.052	0.559
E + I + J	0.407	0.732	0.056	0.571

By comparing  $E$  with  $E + R$  and  $E + I$ , respectively, it could be inferred that combining the mutual information between user node embedding and feature attributes can improve the performance of the algorithms for both types of experiments in this study. For a given user node embedding, maximizing the mutual information between the user node embedding and feature attributes ( $E + I$ ) worked better than minimizing the feature attribute reconstruction error ( $E + R$ ). As shown in Table 3, in individual cases, maximizing the mutual information between user node embedding and feature attributes ( $E + I$ ) was better than optimizing the joint supervised signal ( $E + I + J$ ), but the advantage was not significant enough to prove that the joint supervised signal ( $E + I + J$ ) can further optimize the results of user segmentation. Finally, combined with the experimental data shown in Table 2, the representation fusion proposed in this study reasonably combined different levels of information in multiple heterogeneous networks of OIC, showing that our framework has good robustness and generalization ability with small variance.

The NMI values and attention weights of the layers in the OIC multi-heterogeneous network are shown in Figure 4. The network layers with higher NMI values also had higher attention weights for their corresponding layers, which proves that the method proposed in this study is relatively stable and effective in using representational fusion.

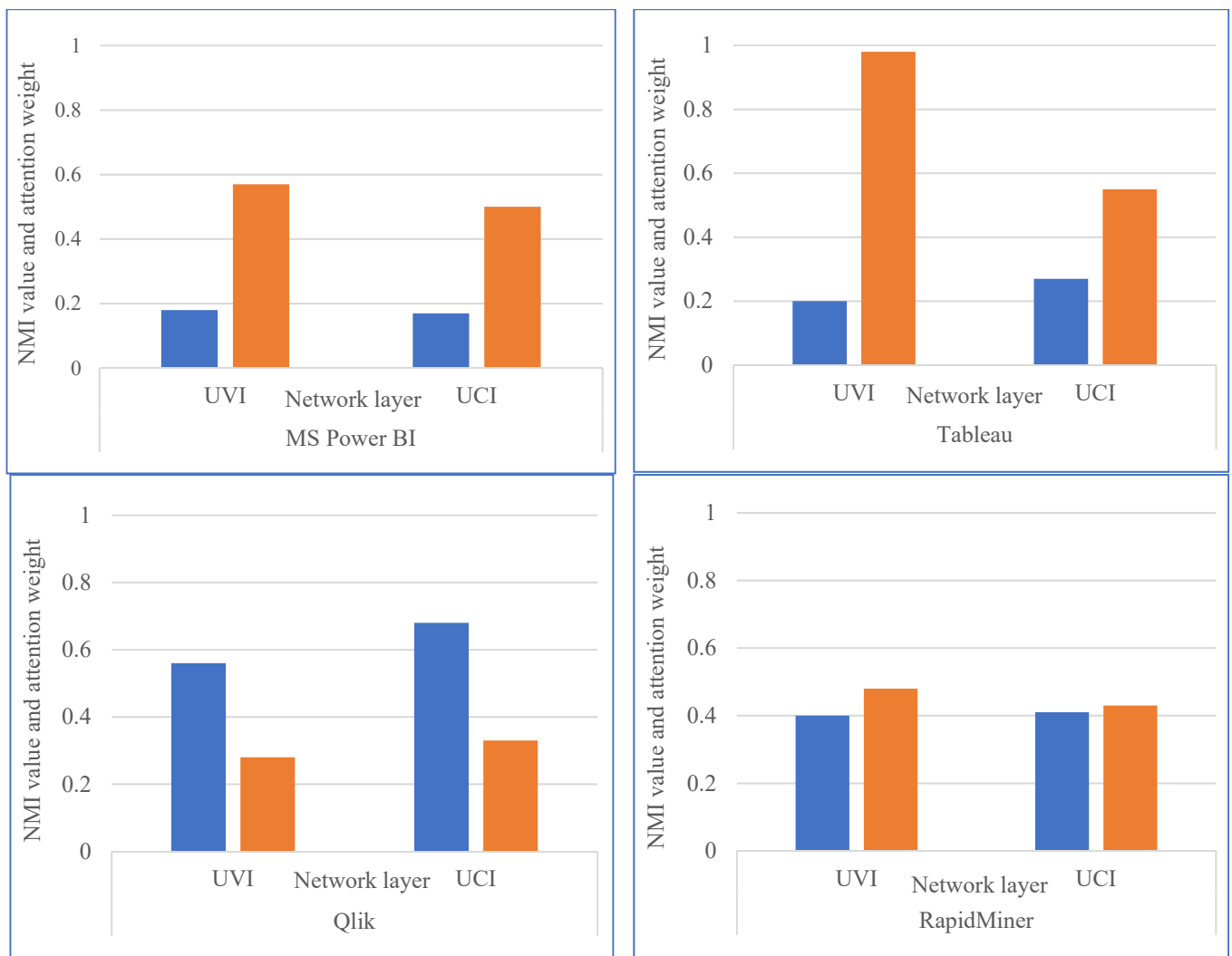


Figure 4. NMI values and attention weights for each layer of the network.

## 6. Limitations and Directions for Future Research

Although the method proposed in this study exhibits better user segmentation performance in OICs than several state-of-the-art baseline methods, the results are limited because the effects of temporal information contained in the heterogeneous OIC and noise points in the network are not considered throughout the study. In future work, relevant denoising techniques can be used to filter out noisy points and to retain the attributes of valid nodes in the real community by transforming the network node attributes and other relevant techniques to filter out nodes that do not meet the basic requirements. This study also has limitations in using data and methods to identify dynamic implicit social relationships in heterogeneous OICs, thus leaving several unexplored areas that provide important directions for future research. First, OICs are dynamic in nature, and future work could focus on recommendation algorithms in dynamic OICs, i.e., introducing temporal factors to describe dynamic user interests and social relationships based on existing recommendation algorithms. Second, in addition to positive relationships between users, there are also negative relationships in OICs, such as distrust or dislike, which is valuable information that has been less studied in current research or other studies. In addition, many real community structures have overlapping components. Therefore, in some cases, it makes more sense to segment overlapping communities. This approach can be considered by preprocessing the network and then combining it with existing overlapping community segmentation algorithms.

Our method, which is limited to matrix operations, is not efficient on large-scale graphs. Therefore, our method is suitable for small-scale graphs with fuzzy community structures. In practice, a more efficient method, such as the Louvain algorithm [59], can be used for initial partitioning, which can subsequently be utilized for finer-grained partitioning. This approach is currently only applicable to undirected networks; however, most networks include directional features. In the future, this method will be investigated to be extended to directed networks. For example, converting directed networks into undirected networks for computation or processing asymmetric connection matrices to meet the requirement of generating symmetric doubly random matrices can be explored. In addition, many real-world community systems contain overlapping components. Therefore, in some cases, the division of overlapping groups is more important. This approach may be used to preprocess the network before combining it with existing overlapping community partitioning mechanisms.

## 7. Conclusions

This study proposes a user segmentation method based on multilayer information fusion and attention mechanism, aiming to accurately and effectively segment user communities in heterogeneous OICs. The method is based on stratifying the heterogeneous OIC according to different edge types to obtain a layered embedding representation of user nodes. Then, the user fusion embedding representation required for user segmentation is obtained by combining the semantic information between the layers. The objective function of mutual information is used to optimize the relevant parameters of user nodes to obtain the final optimized user fusion embedding representation. Finally, the k-means clustering algorithm is used to obtain the results of user segmentation. Experimental analysis of the proposed method on several OIC datasets of business intelligence and analytics software shows that the method outperforms current state-of-the-art community segmentation methods in node clustering and node similarity search experiments and improves the accuracy of user segmentation.

Our method makes a practical contribution to the knowledge creation and innovation process. To better understand community mechanisms and facilitate effective knowledge transfer in the innovation process, our method suggests that user groups exchange data and acquire knowledge from other groups by segmenting users in an open innovation network, which requires both external access and internal capabilities. Internal capabilities can be improved by improving knowledge sharing and contribution capabilities, while external



access to new knowledge can be improved by connecting with other user segments. The study of social relationships within open innovation communities facilitates the transfer and contribution of knowledge, including exchanges between community members, to stimulate the generation of new ideas. As a result of incorporating heterogeneous networks into our method, it is possible to improve not only the performance of user segmentation in heterogeneous OICs, but also the accuracy of knowledge content and incremental predictions.

From a knowledge management perspective, our method can facilitate early detection of high-quality content and common interests in OICs. This not only allows learning from participants of the same group but is also suitable for detecting knowledge patterns and common interests of members of the same group. The combination of network structure and graph-centric networking improves the accuracy of the potential value discovery model of community users' ideas, and also provides technical support for the community to target user participation and fully exploit community innovation resources. Thus, the positioning and relationships of users in the cluster are important for improving the efficiency of the community.

In addition, the communication structure is crucial to the success of many OIC schemes. In this study, we investigate and solve the problem of node neighborhood propagation range constraints in heterogeneous OICs. Our technique is an unsupervised approach with an end-to-end structure capable of performing several downstream tasks (i.e., node classification, similarity search, and node clustering). We build a model that accommodates node neighborhood information at a local scale, while capturing global neighborhood information as well. Additionally, our method removes a portion of edges to increase the unpredictability and diversity of graph connectivity, making the model more resilient and generalizable.

As most of current OICs and platforms support vector computing, using this method to identify the connectivity between nodes in open innovation applications can further enable significant improvements in the quality and efficiency of community operations and knowledge contributions, leading to better innovation management and performance. In our future work, we will investigate more effective dissemination strategies for more complex practical applications, such as knowledge transfer and management embedded in different kinds of community structures.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data supporting reported results can be found in Microsoft Power BI community (<https://community.powerbi.com>), Tableau community (<https://community.tableau.com/s/>), Qlik community (<https://community.qlik.com/>), and RapidMiner community (<https://community.rapidminer.com/>). All Communities were accessed on 1 September 2022.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Daradkeh, M. The Relationship Between Persuasion Cues and Idea Adoption in Virtual Crowdsourcing Communities: Evidence from a Business Analytics Community. *Int. J. Knowl. Manag.* **2022**, *18*, 1–34. [CrossRef]
2. Chesbrough, H. *Open Innovation the New Imperative for Creating and Profiting from Technology*; Harvard Business School Press: Boston, MA, USA, 2003.
3. Daradkeh, M. The Influence of Sentiment Orientation in Open Innovation Communities: Empirical Evidence from a Business Analytics Community. *J. Inf. Knowl. Manag.* **2021**, *20*, 2150029. [CrossRef]
4. Ober, J. Open Innovation in the ICT Industry: Substantiation from Poland. *J. Open Innov. Technol. Mark. Complex.* **2022**, *8*, 158. [CrossRef]
5. Huang, S.; Chen, J.; Wang, Y.; Ning, L.; Sutherland, D.; Zhou, Z.; Zhou, Y. External heterogeneity and its impact on open innovation performance. *Technol. Anal. Strat. Manag.* **2015**, *27*, 182–197. [CrossRef]

6. Chesbrough, H. *Open Innovation Results: Going Beyond the Hype and Getting Down to Business*; Oxford University Press: Oxford, UK, 2019.
7. Marullo, C.; Minin, A.D.; Martelli, I.; Piccaluga, A. Solving the ‘heterogeneity puzzle’: A comparative look at SMEs growth determinants in open and closed innovation patterns. *Int. J. Entrep. Innov. Manag.* **2020**, *24*, 443–464. [[CrossRef](#)]
8. Capasso, M.; Rybalka, M. Innovation Pattern Heterogeneity: Data-Driven Retrieval of Firms’ Approaches to Innovation. *Businesses* **2022**, *2*, 54–81. [[CrossRef](#)]
9. Papa, A.; Mazzucchelli, A.; Ballestra, L.V.; Usai, A. The open innovation journey along heterogeneous modes of knowledge-intensive marketing collaborations: A cross-sectional study of innovative firms in Europe. *Int. Mark. Rev.* **2022**, *39*, 602–625. [[CrossRef](#)]
10. Saebi, T.; Foss, N. Business models for open innovation: Matching heterogeneous open innovation strategies with business model dimensions. *Eur. Manag. J.* **2015**, *33*, 201–213. [[CrossRef](#)]
11. Turoń, K. Open Innovation Business Model as an Opportunity to Enhance the Development of Sustainable Shared Mobility Industry. *J. Open Innov. Technol. Mark. Complex.* **2022**, *8*, 37. [[CrossRef](#)]
12. Muninger, M.; Mahr, D.; Hammedi, W. Social media use: A review of innovation management practices. *J. Bus. Res.* **2022**, *143*, 140–156. [[CrossRef](#)]
13. Fursov, K.; Linton, J. Social innovation: Integrating product and user innovation. *Technol. Forecast. Soc. Chang.* **2022**, *174*, 121224. [[CrossRef](#)]
14. Bachmann, P.; Frutos-Bencze, D. R&D and innovation efforts during the COVID-19 pandemic: The role of universities. *J. Innov. Knowl.* **2022**, *7*, 100238.
15. Urbinati, A.; Manelli, L.; Frattini, F.; Bogers, M. The digital transformation of the innovation process: Orchestration mechanisms and future research directions. *Innovation* **2022**, *24*, 65–85. [[CrossRef](#)]
16. Xiong, B.; Lim, E.T.K.; Tan, C.-W.; Zhao, Z.; Yu, Y. Towards an evolutionary view of innovation diffusion in open innovation ecosystems. *Ind. Manag. Data Syst.* **2022**, *122*, 1757–1786. [[CrossRef](#)]
17. Ferdinand, J.; Meyer, U. The social dynamics of heterogeneous innovation ecosystems: Effects of openness on community—Firm relations. *Int. J. Eng. Bus. Manag.* **2017**, *9*, 1847979017721617. [[CrossRef](#)]
18. Chen, Y.; Hu, Y.; Li, K.; Yeo, C.L.; Li, K. Approximate personalized propagation for unsupervised embedding in heterogeneous graphs. *Inf. Sci.* **2022**, *600*, 287–300. [[CrossRef](#)]
19. Brodny, J.; Tutak, M. Digitalization of Small and Medium-Sized Enterprises and Economic Growth: Evidence for the EU-27 Countries. *J. Open Innov. Technol. Mark. Complex.* **2022**, *8*, 67. [[CrossRef](#)]
20. Ge, J.; Shi, L.; Liu, L.; Shi, H. Intelligent Link Prediction Management Based on Community Discovery and User Behavior Preference in Online Social Networks. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 3860083. [[CrossRef](#)]
21. Jia, J.; Liu, P.; Du, X.; Zhang, Y. Multilayer Social Network Overlapping Community Detection Algorithm Based on Trust Relationship. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 9268039. [[CrossRef](#)]
22. Lu, Q.; Chesbrough, H. Measuring open innovation practices through topic modelling: Revisiting their impact on firm financial performance. *Technovation* **2022**, *114*, 102434. [[CrossRef](#)]
23. Tang, T.; Fisher, G.; Qualls, W. The effects of inbound open innovation, outbound open innovation, and team role diversity on open source software project performance. *Ind. Mark. Manag.* **2021**, *94*, 216–228. [[CrossRef](#)]
24. Weiss-Lehman, C.P.; Werner, C.M.; Bowler, C.H.; Hallett, L.M.; Mayfield, M.M.; Godoy, O.; Aoyomana, L.; Barabás, G.; Chu, C.; Ladouceur, E.; et al. Disentangling key species interactions in diverse and heterogeneous communities: A Bayesian sparse modelling approach. *Ecol. Lett.* **2022**, *25*, 1263–1276. [[CrossRef](#)] [[PubMed](#)]
25. Yuana, R.; Prasetyo, E.A.; Syarif, R.; Arkeman, Y.; Suroso, A.I. System Dynamic and Simulation of Business Model Innovation in Digital Companies: An Open Innovation Approach. *J. Open Innov. Technol. Mark. Complex.* **2021**, *7*, 219. [[CrossRef](#)]
26. Zhang, S.; Tong, H.; Xu, J.; Maciejewski, R. Graph convolutional networks: A comprehensive review. *Comput. Soc. Netw.* **2019**, *6*, 11. [[CrossRef](#)]
27. Perozzi, B.; Al-Rfou, R.; Skiena, S. DeepWalk: Online learning of social representations. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24 August 2014; pp. 701–710.
28. Dong, Y.; Chawla, N.; Swami, A. metapath2vec: Scalable Representation Learning for Heterogeneous Networks. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; pp. 135–144.
29. Jiang, B.; Zhang, Z.; Lin, D.; Tang, J.; Luo, B. Semi-Supervised Learning with Graph Learning-Convolutional Networks. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.
30. Weng, L.; Zhang, Q.; Lin, Z.; Wu, L. Harnessing heterogeneous social networks for better recommendations: A grey relational analysis approach. *Expert Syst. Appl.* **2021**, *174*, 114771. [[CrossRef](#)]
31. Zhou, Z.; Hu, Y.; Zhang, Y.; Chen, J.; Cai, H. Multiview Deep Graph Infomax to Achieve Unsupervised Graph Embedding. In *IEEE Transactions on Cybernetics*; IEEE: Piscataway, NJ, USA, 2022; pp. 1–11.
32. Zhu, X. Incorporation of sticky information and product diversification into static game of open innovation. *Int. J. Innov. Stud.* **2022**, *6*, 11–25. [[CrossRef](#)]

33. Chesbrough, H.; Heaton, S.; Mei, L. Open innovation with Chinese characteristics: A dynamic capabilities perspective. *R D Manag.* **2021**, *51*, 247–259. [[CrossRef](#)]
34. Robaeyst, B.; Baccarne, B.; Duthoo, W.; Schuurman, D. The City as an Experimental Environment: The Identification, Selection, and Activation of Distributed Knowledge in Regional Open Innovation Ecosystems. *Sustainability* **2021**, *13*, 6954. [[CrossRef](#)]
35. Daradkeh, M. Innovation in Business Intelligence Systems: The Relationship Between Innovation Crowdsourcing Mechanisms and Innovation Performance. *Int. J. Inf. Syst. Serv. Sec.* **2022**, *14*, 1–25. [[CrossRef](#)]
36. Abbate, T.; Codini, A.; Aquilani, B.; Vrontis, D. From Knowledge Ecosystems to Capabilities Ecosystems: When Open Innovation Digital Platforms Lead to Value Co-creation. *J. Knowl. Econ.* **2022**, *13*, 290–304. [[CrossRef](#)]
37. Jokar, E.; Mosleh, M.; Kheyrandish, M. Discovering community structure in social networks based on the synergy of label propagation and simulated annealing. *Multimed. Tools Appl.* **2022**, *81*, 21449–21470. [[CrossRef](#)]
38. Xu, X.; Hu, N.; Li, T.; Trovati, M.; Palmieri, F.; Kontonatsios, G.; Castiglione, A. Distributed temporal link prediction algorithm based on label propagation. *Future Gener. Comput. Syst.* **2019**, *93*, 627–636. [[CrossRef](#)]
39. Bai, L.; Cheng, X.; Liang, J.; Guo, Y. Fast graph clustering with a new description model for community detection. *Inf. Sci.* **2017**, *388–389*, 37–47. [[CrossRef](#)]
40. Hammoud, Z.; Kramer, F. Multilayer networks: Aspects, implementations, and application in biomedicine. *Big Data Anal.* **2020**, *5*, 2. [[CrossRef](#)]
41. He, T.; Bai, L.; Ong, Y. Vicinal Vertex Allocation for Matrix Factorization in Networks. *IEEE Trans. Cybern.* **2022**, *52*, 8047–8060. [[CrossRef](#)]
42. Gholami, M.; Sheikhhahmadi, A.; Khamforoosh, K.; Jalili, M. Overlapping community detection in networks based on Neutrosophic theory. *Phys. A Stat. Mech. Appl.* **2022**, *598*, 127359. [[CrossRef](#)]
43. Su, Z.; Lin, S.; Ai, J.; Li, H. Rating Prediction in Recommender Systems Based on User Behavior Probability and Complex Network Modeling. *IEEE Access* **2021**, *9*, 30739–30749. [[CrossRef](#)]
44. Daneshvar, H.; Ravanmehr, R. A social hybrid recommendation system using LSTM and CNN. *Concurr. Comput. Pract. Exp.* **2022**, *34*, e7015. [[CrossRef](#)]
45. Liu, H.; He, L.; Zhang, F.; Wang, Z.; Gao, C. Dynamic community detection over evolving networks based on the optimized deep graph infomax. *Chaos Interdiscip. J. Nonlinear Sci.* **2022**, *32*, 053119. [[CrossRef](#)]
46. Ahmadian, S.; Joorabloo, N.; Jalili, M.; Ren, Y.; Meghdadi, M.; Afsharchi, M. A social recommender system based on reliable implicit relationships. *Knowl. Based Syst.* **2020**, *192*, 105371. [[CrossRef](#)]
47. Huang, M.; Jiang, Q.; Qu, Q.; Chen, L.; Chen, H. Information fusion oriented heterogeneous social network for friend recommendation via community detection. *Appl. Soft Comput.* **2022**, *114*, 108103. [[CrossRef](#)]
48. Awati, C.; Shirgave, S. The State of the Art Techniques in Recommendation Systems. In *Applied Computational Technologies*; Springer: Singapore, 2022.
49. Schutera, M.; Rettenberger, L.; Pylatiuk, C.; Reischl, M. Methods for the frugal labeler: Multi-class semantic segmentation on heterogeneous labels. *PLoS ONE* **2022**, *17*, e0263656. [[CrossRef](#)]
50. Jia, X.; Shang, J.; Liu, D.; Zhang, H.; Ni, W. HeDAN: Heterogeneous diffusion attention network for popularity prediction of online content. *Knowl. Based Syst.* **2022**, *254*, 109659. [[CrossRef](#)]
51. Xu, X.; Chen, C.; Mendes, J. Quantifying dissimilarities between heterogeneous networks with community structure. *Phys. A Stat. Mech. Appl.* **2022**, *588*, 126574. [[CrossRef](#)]
52. Han, Z.; Huang, Q.; Zhang, J.; Huang, C.; Wang, H.; Huang, X. GA-GWNN: Detecting anomalies of online learners by granular computing and graph wavelet convolutional neural network. *Appl. Intell.* **2022**, *52*, 13162–13183. [[CrossRef](#)]
53. Wang, X.; Bo, D.; Shi, C.; Fan, S.; Ye, Y.; Yu, P.S. A Survey on Heterogeneous Graph Embedding: Methods, Techniques, Applications and Sources. *arXiv* **2022**, arXiv:2011.14867. [[CrossRef](#)]
54. Liu, X.; Tang, J. Network representation learning: A macro and micro view. *AI Open* **2021**, *2*, 43–64. [[CrossRef](#)]
55. Zhang, Z.; Chen, C.; Chang, Y.; Hu, W.; Xing, X.; Zhou, Y.; Zheng, Z. SHNE: Semantics and Homophily Preserving Network Embedding. In *IEEE Transactions on Neural Networks and Learning Systems*; IEEE: Piscataway, NJ, USA, 2021; pp. 1–12.
56. Giraudo, M.; Sacerdote, L.; Sirovich, R. Non-Parametric Estimation of Mutual Information through the Entropy of the Linkage. *Entropy* **2013**, *15*, 5154–5177. [[CrossRef](#)]
57. Chowdhury, K.; Chaudhuri, D.; Pal, A. An entropy-based initialization method of K-means clustering on the optimal number of clusters. *Neural Comput. Appl.* **2021**, *33*, 6965–6982. [[CrossRef](#)]
58. Daradkeh, M. Exploring the Usefulness of User-Generated Content for Business Intelligence in Innovation: Empirical Evidence from an Online Open Innovation Community. *Int. J. Enterp. Inf. Syst.* **2021**, *17*, 44–70. [[CrossRef](#)]
59. Hu, J.; Wang, Z.; Chen, J.; Dai, Y. A community partitioning algorithm based on network enhancement. *Connect. Sci.* **2021**, *33*, 42–61. [[CrossRef](#)]
60. Qin, J.; Zeng, X.; Wu, S.; Zou, Y. Feature recommendation strategy for graph convolutional network. *Connect. Sci.* **2022**, *34*, 1697–1718. [[CrossRef](#)]
61. Berahmand, K.; Nasiri, E.; Rostami, M.; Forouzandeh, S. A modified DeepWalk method for link prediction in attributed social network. *Computing* **2021**, *103*, 2227–2249. [[CrossRef](#)]

62. Grover, A.; Leskovec, J. node2vec: Scalable Feature Learning for Networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 855–864.
63. Zhang, T.; Xiong, Y.; Zhang, J.; Zhang, Y.; Jiao, Y.; Zhu, Y. CommDGI: Community Detection Oriented Deep Graph Infomax. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, Virtual Event, Ireland, 19–23 October 2020; pp. 1843–1852.
64. Li, L.; Jin, L.; Zhang, Z.; Liu, Q.; Sun, X.; Wang, H. Graph Convolution Over Multiple Latent Context-Aware Graph Structures for Event Detection. *IEEE Access* **2020**, *8*, 171435–171446. [[CrossRef](#)]