IJE TRANSACTIONS B: Applications Vol. 36 No. 11, (November 2023) 2052-2062

International Journal of Engineering

Journal Homepage: www.ije.ir

A Hybrid Approach to Detect Researchers' Communities Based on Deep Learning and Game Theory

A. Torkaman[a], K. Badie[b], A. Salajegheh*[a], M. H. Bokaei[c], S. F. Fatemi Ardestani[d]

[a] *Department of Computer, South Tehran Branch, Islamic Azad University, Tehran, Iran*
[b] *E-Services and E-Content Research Group, IT Research Faculty, ICT Research Institute, Tehran, Iran*
[c] *Department of Information Technology, ICT Research Institute, Tehran, Iran*
[d] *Sharif University of Technology, Tehran, Iran*

PAPER INFO

Paper history:
Received 18 June 2023
Received in revised form 24 July 2023
Accepted 25 July 2023

Keywords:
Complex Networks
Community Detection
Deep Learning
Cooperative Game
Non-Cooperative Game

ABSTRACT

Today, with the proliferation of complex networks and their large amounts of data, researchers have great concerns about the accurate community detection methods. The difficulty in analyzing these networks stems from their enormous size and the complex relationships among the members of the networks. It is difficult to analyze the deep relationships and mechanisms by just looking at the whole. Traditional methods have some problems and limitations when analyzing these networks such as feature extraction, high reliance on the initial phase settings, computational complexity, neglect of network relationships and content. From the perspective of relationships and interactions between individuals, the environment of complex networks can be compared to a game in which nodes acting as players or agents may join or leave a community based on similar structural or semantic characteristics. Consequently, there is a strong tendency to use cooperative and non-cooperative games to detect communities. Moreover, the amalgamation of deep learning techniques and game theory has recently been proven to be highly effective in extracting communities. Deep learning techniques have demonstrated enhanced capability in feature engineering and automate the process. In this study, the authors make effort to detect rational and accurate communities based on structural and content features with the help of traditional approaches, deep learning, as well as cooperative and non-cooperative games. The efficiency of this study is demonstrated by experimental findings on real datasets, and confirming that it is able enough to identify those communities that are more meaningful.

doi*: 10.5829/ije.2023.36.11b.10*

1. INTRODUCTION

One of the most crucial research area of study in complex networks is community detection. This has motivated many researchers over the years to find node groups based on modular patterns .

The ability to detect communities gives us further intuition into how groups function and how they form. Nodes within a community share similar characteristics and interests with one another and are more closely connected than other nodes within the network.

Various studies have provided different perspectives on communities extraction, such as partitioning approaches, hierarchical methods, edge removal methods, as well as factorization-based and modularity-based approaches [1-3]. However, these scenarios may work well in some situations, as the issue of detecting community is inherently challenging and involves multiple factors, it is better to look at this issue from a different aspect.

In citation networks, communities are formed solely based on individuals self-interest. It is only in the interest of individuals to decide on their membership [4]. In one hand, we can conceive about a cooperative environment in which individuals connect with each other, form communities, and seek to promote the utility of the group, so there is also a kind of coordination between them.

*Corresponding Author Email: a_salajegheh@azad.ac.ir (A. Salajegheh)

Please cite this article as: A. Torkaman, K. Badie, A. Salajegheh, M. H. Bokaei, S. F. Fatemi Ardestani, A Hybrid Approach to Detect Researchers' Communities Based on Deep Learning and Game Theory, ***International Journal of Engineering, Transactions B: Applications***, Vol. 36, No. 11, (2023), 2052-2062

On the other hand, we can also imagine a competitive environment where agents compete to join or leave their communities and increase their profits.

Arguably, understanding these relationships requires rationality. By applying the theoretical economic principles of game theory, makes it easy to analyze these relationships.

In order to produce logical and ideal answers in challenging circumstances, game theory is a highly helpful mathematical instrument for analyzing strategic conditions and modelling the competition and collaboration between decision-makers [5]. Game theory is generally divided into cooperative and non-cooperative types. Games in which emphasize member cooperation is referred to cooperative game [6], in which each player attempts to increase the utility of the coalition [7]. In contrast, in non-cooperative games, players ignore the gains of the group and focus on increasing their individual utility.

Therefore, we address the idea of using cooperative and non-cooperative games to obtain more satisfying and trustworthy communities.

Recently, deep learning techniques along with game theory have proven to be extremely useful in extracting communities. Deep learning can provide features that are more informative and open up new perspectives in solving the community detection problems in large-scale networks.

In this respect, we utilized three well-known deep learning algorithms to obtain instructive characteristics. First, we learnt the embedding vectors from the network structure through the DeepWalk method [8]. Then, to extract the content features, we used LSTM [9] and Doc2vec [10] algorithms. After preparing the features, to identify the primary clusters, we applied a popular partitioning technique such as K-means to divide the network and provide the initial clusters. This helps reduce computational complexity. Rather than trying to initiate clustering through an agglomerative hierarchical approach.

Then, to stabilize the initial clusters and decrease them, we utilized the advantages of cooperative games. Meanwhile, we need to make sure that the nodes are properly assigned to the communities. So, we considered the privileges of non-cooperative games (Figure 1).

As the cooperative game is applied based on k-means clustering results instead of singleton clustering, the complexity is reduced. Additionally, since the non-cooperative game applied on the stabilized clusters extracted by cooperative game, so nodes as selfish players can only be compared with the significant and important nodes who have a high degree in their communities. As a result, there are fewer comparisons conducted as well as fewer nodes planning to leave their community due to improving their utilities.

Totally, the suggested method is more efficient and the computing cost noticeably reduced.

The main contributions of this paper are as follows.

- In this research, we used three well-known deep learning algorithms to obtain instructive features. Both structural and content features have been extracted through deep learning to help identify more meaningful communities. The proposed model can deal with long-term dependencies and solve the vanishing gradient problem.
- In order to reduce the computational complexity, rather than trying to initiate clustering by using an agglomerative hierarchical approach, we used a traditional clustering method, k-means, to divide the network and provide the initial clusters.
- Use of traditional clustering methods such as K-means does not guarantee the clusters obtained and may generate more clusters than the actual ones, so by using cooperative game theory we can reduced and stabilized the extracted communities.
- In some cases, there may be a limited number of nodes belonging to different communities, or there may be single nodes that do not belong to a proper community. In such cases, we utilized the benefits of non-cooperative game which helps rationally

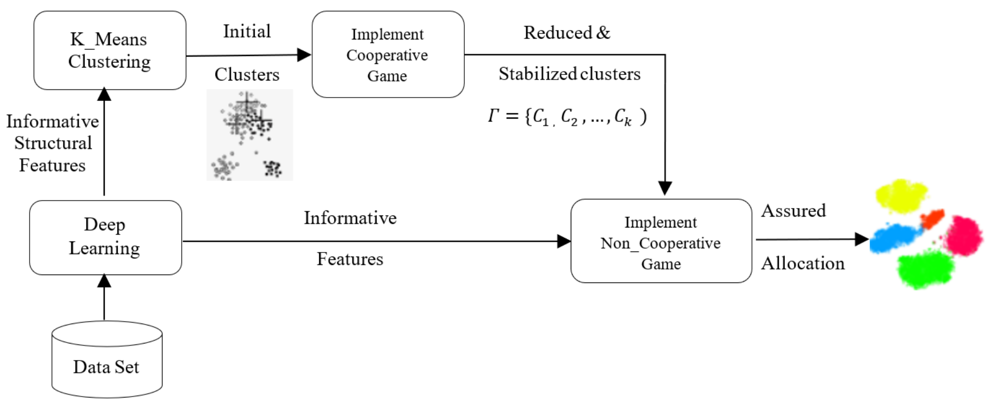

Figure 1. The proposed framework

allocate these nodes to well-established communities.

- Individual nodes which have no connections with the rest of the network can also be assigned to communities according to their content.

The following sections of this essay are structured as follows: Section 2 reviews the literature on several approaches to the challenge of community detection. In section 3, the suggested model is mentioned. Section 4 discusses the analysis of the experimental findings, and section 5 presents concluding remarks and further research.

2. RELATED WORK

A community consists of multiple components that are in close proximity to each other only within their respective groups, in contrast to the rest of the network. Individuals that reside in the same community share similar characteristics, such as interests, social links, locations, occupations, etc. [11, 12].

The nodes belonging to the same community typically have similar responsibilities and/or functions [13].

Community detection is one of the most exciting research areas, which has caught the interest of numerous researchers in a variety of fields of study, including biology, statistics, and computer science [14]. Community detection is typically an NP-complete issue [13, 14].

Several studies published in the literature attempt to extract high-quality communities. Some of them used the graph partitioning techniques such as K-means to detect communities [15-19].

However, these methods have significant drawbacks like weak cluster descriptors and high sensitivity to initialization. Hierarchical clustering based on agglomerative or divisive methods is another aspect widely used in literatures [20-23]. Some researchers focused on improving quality metrics such as modularity for obtaining a high-quality community structure [24].

In recent years, methods based on network representations by using deep learning have become popular for community detection. Some of them respected the structure perspective [8, 25-27]. Salehi and Pouyan [11] proposed a model for detecting communities within social networks based on deep learning. In this method, a nonlinear embedding of the original graph is fed to stacked auto-encoders to train. Then a clustering algorithm is employed to extract communities. However, These methods work well, but they only consider the structural information and ignore node content information.

Other methods, incorporate the node content information into network representation [28-30].

These approaches work logically. However, it is beneficial to use methods that are based on both structural and content information to detect communities.

Over the past two decades, various studies have proposed game-theoretic approaches to identify communities.

In fact, community detection can be likened to a game in which each node makes rational decisions about which community to join in order to maximize its score. Additionally, community members try to increase the utility of group.

Many researches addressed the challenge of community detection by using the non-cooperative game, while others use cooperative one. In line with the cooperative game, Mcsweeney et al. [31] treated each node as a player in a hedonic game, which aims to create an stable community structure.

The Shapley value was recommended by Zhou et al. [32] to identify communities within a specific social network. They also suggested a coalitional game for detecting communities based on the node structure in 2015.

Each node was envisioned by Hajibagheri et al. [33] as a logical being attempting to maximize the Shapley value. Avrachenkov and his colleagues [34] suggested two cooperative game theory methods based on hedonic and Myerson value games for detecting communities. Both methods extract communities with varying resolutions.

Nodes were taken into account by Zhou et al. [35] as players attempted to increase the utility of their coalitions by taking part in a cooperative games. In this study, an edge weight computation was proposed to determine the Shapley value for coalitions and nodes.

According to Chen et al. [36], on the non-cooperative side, agent's utility is computed as a gain and loss function based on modularity and community membership, respectively. Therefore, the game's local equilibrium reveals community organization at the end. Furthermore, Narayanam and Narahari [37] believed that the utility of each vertex is a linear function, imagining each node as an agent wanting to join a community. In this study, community stability is ensured via Nash stability.

A methodology based on the iterative game have been considered by Alvari et al. [38] for detecting communities in complex networks. They considered nodes as logical players who enter the game to increase their utility.

A weighted potential game was developed by Havvaei and Deo [39] to demonstrate community structure. When a community reaches the Nash equilibrium point, it stabilizes. Co-game is a game-theoretical method for identifying communities in real-world networks, as described by Zhao et al. [40]. This technique combines individual games and equilibrium to create finer-grained partitions in the detection process.

A game-theoretical algorithm for detecting communities in online complex networks was developed by Vincenzo et al. [41]. They modeled the process of community formation as a game, in which each node as a player aiming to maximize its goals. They used a game theory approach to simulate how communities form. Each node is regarded as a player trying to maximize its utility.

SIMGT [42] is a useful method for identifying communities, which assumes nodes as self-interested players participating in a non-cooperative game. To update players identities, they used a stochastic gradient ascent.

Zhou et val. [43] proposed a novel method for detecting communities based on both cooperative and non-cooperative games. This method imagined nodes as players in coalitional form games who want to increase the utility of the group, meanwhile playing non-cooperative games to increase their own utility.

Similar to the hierarchical agglomerative method, this approach considers a cooperative game in the initial phase, where nodes or agents are clustered as singletons, and coalitions with the highest utility value are combined into larger coalitions until high-quality coalitions are attained. This method, like other agglomeration approaches typically has considerable computational cost for large data sets. Therefore, it is recommended to integrate the first phase with other clustering techniques.

In this regard, Torkaman et al. [44] proposed a Four-Stage Algorithm (FSA), which find the important central nodes, propagate labels, and identify initial communities to solve this problem. However, this method focuses only on structural information and omits content information.

Therefore, in this study, we proposed a new community detection model based on both structural and content features, using a traditional clustering method to reduce the initial computational cost, and integrate cooperative and non-cooperative games to provide reliable and stable communities.

3. PROPOSED METHOD

3. 1. Preliminary A citation network is a type of complex network that may include various papers, books, linked by co-citation relationships. A key issue in network analysis is how to represent these networks. Assume a network G=(V,E,D) and a set of vertices V={v_1, v_2 , v_3, …,v_n}, n is the number of vertices. $E = \{e_{ij}\}_{i,j=1}^{n}$ the set of edges, and the edge among v_i and v_j is encoded as $\boldsymbol{e_{ij}}$. D is the set of textual data which relates to each node of v_i.

The goal of the network-embedding problem is to develop the mapping function: $f: V \rightarrow \mathbb{R}^d$ which maps each node into a low-dimensional space and extracts the network's structural and content characteristics. Nodes in this representation space that have similar structure or content are located close to one another.

3. 1. 1. Community Detection Community detection is an operation to detect M communities; C= {C1, C2,…, CM} , so M ≪N and $\cup_{m=1}^{M} C_m = V$. if $C_i \neq C_j$ for any subset of V, then nodes can only join one community and are referred to as non-overlapping communities. If it can join more than one community, it entitled as overlapping communities [45].

3. 1. 2. Game Theory Background Game theory is a mathematical tool that focuses on decision-making problems between two or more entities engaged in strategic scenarios where one player's decisions affect the other players' payoffs [46].

The interaction between vertices in a complex network may be compared to a game in which a node acts like a player and seeks to join or leave the community depending on its utility in a target community.

Let u_i be the utility of vertex $i \in V$. For everyC_i, $u_i\,(C_i)$ is the utility function of i by existing in a community C_i. Every vertex seeks to become a member of a community and increase its own utility. It should be emphasized any node's utility is depends on the community to which is it belongs.

3. 2. Our Proposed Model This study utilizes deep learning approaches and game theory to find established and accurate communities, as previously described.

Figure 1 shows our suggested framework. First, according to our previous work [44], we extract informative features by using three popular deep learning approaches.

The architecture of the proposed representation is shown in Figure 2. It projects each node into a low dimensional region to capture the network's structural and content properties [44].

As shown in Figure 2, we learn embedding vectors from network structure by DeepWalk [8] method, to provide deep structural features to the K-means clustering algorithm. Then, to extract the content features, we used the concatenated vectors of both Long Short-Term Memory (LSTM) [9] and Doc2vec [10] algorithms, to enhance each other to extract the context sequence from the paper's titles or abstracts more accurately. Therefore, in this paper, we applied the same loss function used by Torkaman et al. [44] to extract the structural and content properties of the nodes:

$$\mathcal{L} = \mathcal{L}_{Doc2Vec} + \mathcal{L}_{LSTM} + \mathcal{L}_{Deepwalk} \quad (1)$$

Afterward, in order to partition the network and find initial communities, we applied a common partitioning

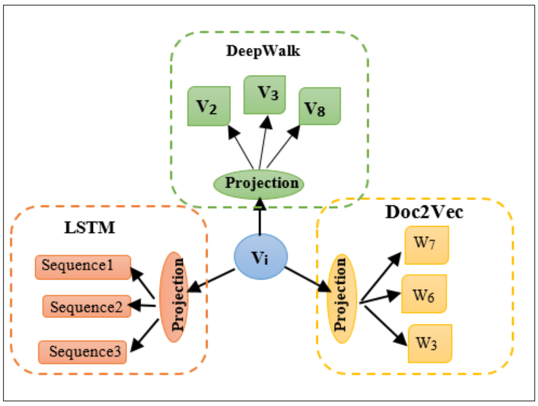

Figure 2. Architecture of our network representation

technique such as the k-means algorithm [47]. This helps reduce computational complexity rather than trying to initiate clustering by using an agglomerative hierarchical approach.

The elbow approach is used to find the optimum value of K [48], which is one of the most popular methods for choosing the ideal value of K.

Therefore, the initial clusters B=$\{B_1, B_2, \dots, B_k\}$, or coalition structure is prepared. In many instances, the number of initially extracted communities is numerous or located far from the actual desired communities that they become unsustainable. Thus, we used the advantages of cooperative (coalitional form) and non-cooperative games to decrease the number of these communities, make them similar to the actual ones and precisely map nodes to communities.

Implementing Cooperative Game: Nodes in a citation network are assumed to act as rational agents striving to form communities (coalitions) and increase the utility of groups. Fewer nodes communities are merged with bigger ones until the merge operation no longer improves the utility of the merged coalitions. In this case, neither coalition intends to cooperate with the other because the game has reached equilibrium. The game starts with an initial cluster from the K-Means algorithm.

Given B_i is a coalition of G= <V, E>, which is achieved by k-means method.

Definition 1: The utility function of coalition, $u(B_i)$ of B, is based on the function was described in our previous work [43]:

|E| is the total number of edges in G, e (B_i) is the number of edges connecting vertices withinB_i, and D(B_i) is the sum of the degree of the vertices in B_i. In fact, $u(B_i)$ comes from Newman's modularity metric Q [49].

In general, if $\Delta u(B_i, B_{ij}) > 0$ and $\Delta u(B_j, B_{ij}) > 0$ then two coalitions are merged [43]. Communities that recently joined are added to a new list $\Gamma = \{C_1, C_2, \dots, C_n\}$.

The final extracted coalitions are stable if there are no coalition intends to participate in merge operation to increse its utility. That is, if $u(B_i) > u(B_i + B_j)\ \forall\ B_j \neq B_i$, B_i does not want to join B_j and it favors to remain within the past situation. In this way, an equilibrium state of the coalition is achieved.

Implementing Non-Cooperative Games: After reaching a set of stable communities, non-cooperative games are played. Single nodes with connections that might not satisfy with their utilities may not be in their correct coalition.

Each node in this game is viewed as a selfish agent who seeks to join or leave a community from Γ depending on its utility measure. It would leave its existing alliance and join a new one, if joining a coalition would increase its utility.

Since nodes in a citation network have structure and content information, the utility function should be the combination of them, especially in the case of single nodes, which can only be determined based on their content similarity.

Definition 2: The utility function of an individual (node): Let $v \in V$, $C_i \in \Gamma$, the utility function is as follows:

$$u_v(C_i) = \alpha W + \beta \frac{e(v,C_i)}{d(v)} \tag{3}$$

$e(v, C_i)$; the number of edges among v and coalition C_i . $d(v)$ is the degree of v and W is the informative feature vector. α, β are binary value $\alpha = 0$, which means the network only consist of structural value. $\beta = 0$ means v is a singleton node and just the similarity determines the closeness between node v an C_id. $u_v(C_i)$ measures the similarity between $\boldsymbol{v}$ and the targeted community C_i. The greater value of $u_v(C_i)$, indicates more similarity between $\boldsymbol{v}$ and C_i .

Definition 3: (Join & Leave): node v join the community C_i :

$C_i \leftarrow C_i + \{v\}$

If $v \notin C_i$ and $u_v(C_i)$ is the maximum value that v can achieve through joining communities.

Node v leave its community C_n and join community C_i:

$C_n \leftarrow C_n - \{v\}$

if $v \in C_n$ and $u_v(C_n) < u_v(C_i)$.

Finally, when agents have no incentive to leave their own community and join others, a kind of equilibrium has prevailed and communities have reached a stable state.

Cooperative and non-cooperative algorithms described in Algorithms 1 and 2.

$$u(B_i) = \sum_{B_i \in B} \left(\frac{e(B_i)}{|E|} - \left(\frac{D(B_i)}{2.|E|} \right)^2 \right) \tag{2}$$

Algorithm 1 *Cooperative Game*

1: Input: the initial coalitions achieved by k-means algorithm B=$\{B_1, B_2, \dots, B_k\}$

2: Output: Community reduction and stabilization $\Gamma = \{C_1, C_2, \dots, C_n\}$

$\Gamma = \{\}$
for all ($B_i, B_{j \in} B$ and $B_j \neq B_i$) do
  if $\Delta u(B_{ij}) > \Delta u(B_j)$ & $\Delta u(B_j) > 0$ then
    $\Gamma = \{B_{ij}\} - \{B_i\} - \{B_j\}$
  else
    return Γ
  end if
end for
(Continue until no coalition is willing to join the other in order to enhance its utility)

As already mentioned, the initial set of communities (B) is provided by the K-means algorithm. Then, the game initials are between these communities. Given B_i, B_j two communities in B, If the union of these two communities (B_{ij}) has more benefits than either community alone, a join operation takes place and B_{ij} is added to the new list Γ. The algorithm may terminate when no coalition intends to use the join mechanism and improve its utility.

Algorithm 2 *non-Cooperative game*
Input: the cooperative game's reduced and stabilized communities $\Gamma = \{C_1, C_2, \ldots, C_n\}$
Output: Node allocation assurance and ultimate stable community structure $C = \{C_1, C_2, \ldots, C_n\}$
δ {}
for all ($v \in C_i$) do
  $\delta = C - C_i$
  for all ($C_j \in \delta$) do
    if $(\Delta u_v(C_j)) > (\Delta u_v(C_j))$
      $C_j = C_j + \{v\}$
      $C_i = C_i - \{v\}$
    end if
    return C_i, C_j
  end for
end for
(Continue until nodes are not eager to leave their current communities and join new ones.)

All nodes are initially placed in their own community provided by Algorithm 1. Each vertex assesses other communities and determines how useful it is to join them. If the value exceeds the utility, it quits its current coalition and joins the new one. The algorithm will stop when the agents chooses to remain in their current situation rather than join other communities to increase their utility values.

4. EXPERIMENTS

In this section, extensive experiments are conducted to validate the effectiveness of our proposed method. two widely used real-world datasets and six state-of-the-art baselines are adopted for the experiment.

4. 1. Datasets We use actual networks to demonstrate the potential and efficiency of the suggested approach, and observe the experimental results. In this regard, we run the proposed approach on two real citation networks (DBLP [50], Citeseer [51]).

Dblp: Dblp is a well-known citation network containing bibliography data in computer science. In total, it includes 60,744 papers and 52,890 edges and four research areas consist of data mining, database, the artificial intelligent, and computer vision.

Citeseer [51]: This data set is a citation network of computer science publications. It contains 3312 publications and 4,732 edges, each of the papers is labeled as one of six categories, artificial intelligence, agents, database, human–computer interaction, information retrieval, and machine learning.

4. 2. Baseline Learning Algorithms For a comprehensive evaluation, we compare our proposed algorithm with a number of methods from different categories.

K-means [47]: A popular classical shallow partitioning algorithm for clustering, alternately updates the location of the cluster center and the distance of the sample from the cluster center.

Spectral [52]: A classical shallow clustering method based on graph theory, using the node adjacency matrix as the similarity matrix.

Louvain [53]: The Louvain Method is a widely used greedy algorithm for community detection by network modularity maximization.

ARGAE [54]: The Adversarial Regularized Graph Autoencoder (ARGAE) method is a graph clustering method, where a discriminator is utilized to ensure the deep representation calculated by encoder matching a prior distribution.

DAEGC [55]: deep attentional embedded graph clustering (DAEGC), is a graph clustering method utilizing a self- optimizing module to learn a clustering-oriented deep representation.

MGCCN [56]: Multilayer Graph Contrastive Clustering Network (MGCCN), a generic and effective autoencoder framework for multilayer graph clustering.

4. 3. Analysis of Experimental Results We calculated the accuracy(ACC) [57], the Normalized mutual information (NMI) [32], the purity measure [13] and between the extracted community structures, taking into account the ground truth of the datasets as an evaluation metric.

We evaluate the effectiveness of the suggested algorithm based on the above metrics, Table 1 and Figure 3 show the accuracy, purity and NMI values of only k-means after applying the cooperative game and eventually running the non-cooperative game.

All the algorithms were run in Python on a desktop PC with an Intel Core i7 CPU (3.4 GHz) and 8 GB RAM.

According to the results (Table 1), after executing the cooperative game with k-means results, ACC, NMI and purity scores improved due to the cluster merging process and a stable point was reached. K-means method does not work well in this situation due to its restrictions such as weak cluster descriptors and its high degree of sensitivity to initial parameters such as determination of the K values.

As is shown in Figure 3, promising results were obtained by running the non-cooperative game on the results of the cooperative strategy.

This is because each node tries to join a community or leave the current community based on its semantic and structural features. In particular, individual nodes that are not connected to a community can easily increase their utility by joining a community based on semantic features. This situation is obvious because the dblp dataset contains a significant amount of single nodes. Once the equilibrium point is reached, all nodes and communities are in stable state.

For a comprehensive evaluation, we compared our proposed model with different methods in different categories.

As shown in Table 2 and Figure 4, it can be observed that our proposed method achieves competitive performance compared with all the baseline methods according to two clustering metrics, which demonstrates the effectiveness of our method. Specifically, we can make the following interesting observations:

TABLE 1. Accuracy (ACC), Normalized mutual information (NMI) and Purity evaluation metrics after running each method on the Dblp, Citeseer dataset with ground truth

Metric	Purity		NMI		ACC	
	Datasets					
Methods	**Citeseer**	**Dblp**	**Citeseer**	**Dblp**	**Citeseer**	**Dblp**
K_Means	0.618	0.735	0.312	0.431	0.544	0.604
Cooperative Game	0.678	0.739	0.345	0.433	0.612	0.709
Non- Cooperative Game	0.853	0.886	0.597	0.762	0.788	0.907

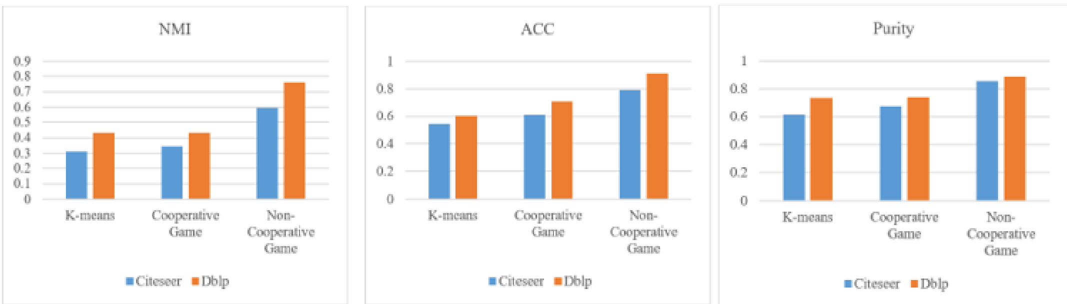

Figure 3. Accuracy (ACC), Normalized mutual information (NMI) and Purity evaluation metrics after running each method on the dblp dataset with ground truth

TABLE 2. Clustering results on Citeseer and Dblp datasets

Metric Methods	NMI		ACC	
	Datasets			
	Citeseer	**Dblp**	**Citeseer**	**Dblp**
K-means	0.312	0.431	0.544	0.604
Spectral	0.056	0.223	0.239	0.402
Louvain	0.409	0.504	0.437	0.513
ARGAE	0.350	0.495	0.573	0.605
DAEGC	0.397	0.561	0.672	0.869
MGCCN	0.455	0.615	0.715	0.830
Ours	0.512	0.762	0.788	0.907

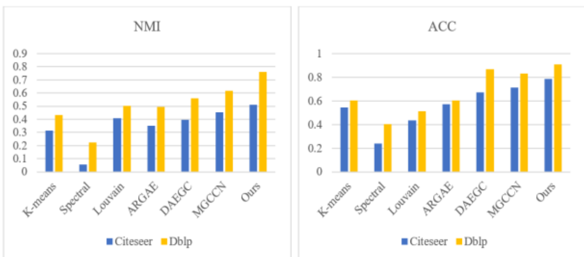

Figure 4. ACC and NMI comparison results on different datasets and methods

- The proposed method and other Graph Convolutional Network (GCN) based methods (DAEGC, MGCCN) show superiority over K-Means, Louvain and Spectral methods, which demonstrates that methods based on both structural and information characteristics, performs better than only using one of them.
- According to the results, our proposed algorithm outperformed other methods in Dblp and Citeseer. In fact, this algorithm yielded higher values of accuracy and NMI than the other existing methods. This is due to utilizing the combination of deep learning and game theory to find established and accurate communities.
- Our method yields a relative increase in NMI values of 35.03% for ARGAE and 26.3% for DAEGC for the Dblp dataset, and the increase is even greater for the Citeseer dataset. These GCN-based approaches use adjacency matrices to represent topological features, but have limitations on large datasets.
- MGCCN is a close competitor to our method, but as we can see, our method has better performance. MGCCN is a generic framework which designed for multi-view graph clustering. MGCCN employs a self-supervised component that iteratively updates the node embedding and clustering, so, there is no guarantee that samples will be assigned to the correct clusters. Therefore, in some cases, "highly confident" nodes are used that act as a soft label to supervise the clustering process. While our method uses both cooperative and non-cooperative methods to solve this problem, and the resulting clusters are reliable and stable.

4. 4. Parameter Sensitivity We have used the same parameter settings that are reported by Torkaman et al. [44] for deep learning part. We set window size b=8, The embedding size is set to k =300 (100 for each proposed deep learning methods), the learning rate= 0.001 and Adam [58] as the optimizer.

Game theory has been used as a tool to achieve more reliable and stable communities. The game parameters are generic and we do not interfere in its settings.

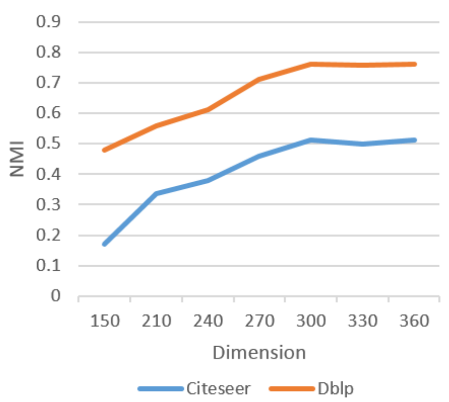

Figure 5. The effect of embedding size on the NMI result

One of the important parameters in the proposed model is the embedding size. the proper size for this vector is 100 for each structural and content vectors. If the size of the vector exceeds this value, the efficiency of the proposed algorithm does not change significantly, only the dimensionality of the problem increases (Figure 5).

4. 5. Network Visualization Finally, we leverage 2D t-SNE projection [59, 60] to visualize the results of the community detection method applied to the dblp dataset.

As shown in Figure 6(a), in the visualization of K_Means, the clusters are not so clear. In Figure 6(b), the cooperative game method outperforms K_Means, but for some classes do not have a clear resolution. In Figure 6(c), after using the non-cooperative approach, the clusters become clearer and almost have a meaningful layout for each community.

5. CONCLUDING REMARKS

In this paper, we proposed a robust and efficient community detection approach that integrates both the topological and content information for community detection. To find the initial clusters, we first used a traditional clustering technique, K-means. Then, to decrease the obtained clusters and fix them, we used a cooperative game, and finally play a non-cooperative game on each node to guarantee a fair and rational allocation of nodes to the established communities.

Experimental findings support the effectiveness of our method, showing how cooperative and non-cooperative game techniques complement each other to identify safer and more stable communities, thereby improving K-means results.

The proposed algorithm shows high performance on medium-sized datasets. However, there are limitations for very large networks with many extracted communities, it takes time to identify the right communities.

Future work may use a different splitting method to split the network and provide initial communities instead of using K-means.

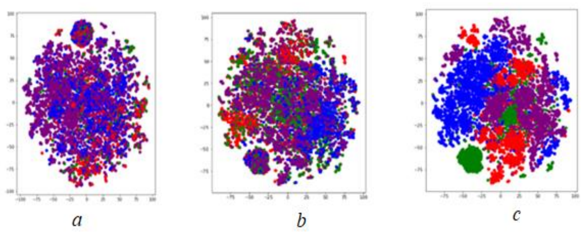

Figure 6. Visualization on the dblp dataset. Colors demonstrate the ground-truth communities

Additionally, various deep learning techniques can be used to extract more advantageous features. The utility functions of the game components, are supposed to be replaced by the other deep learning methods. Finally, this framework can be extended not only to weighted networks, but also to overlapping networks with semantic content.

6. REFERENCES

1. Fortunato, S., "Community detection in graphs", ***Physics Reports***, Vol. 486, No. 3-5, (2010), 75-174. https://doi.org/10.1016/j.physrep.2009.11.002
2. Coscia, M., Giannotti, F. and Pedreschi, D., "A classification for community discovery methods in complex networks", ***Statistical Analysis and Data Mining: The ASA Data Science Journal***, Vol. 4, No. 5, (2011), 512-546. https://doi.org/10.1002/sam.10133
3. Krishna Prasad, M. and Rao Chintalapudi, S., "Mining overlapping communities in real-world networks based on extended modularity gain", ***International Journal of Engineering, Transactions A: Basics***, Vol. 30, No. 4, (2017), 486-492. doi: 10.5829/idosi.ije.2017.30.04a.05.
4. Jonnalagadda, A. and Kuppusamy, L., "A survey on game theoretic models for community detection in social networks", ***Social Network Analysis and Mining***, Vol. 6, (2016), 1-24. doi: 10.1007/s13278-016-0386-1.
5. Nash, J., "Non-cooperative games. The annals of mathematics, second series, volume 54, issue 2", (1951).
6. Zlotkin, G. and Rosenschein, J.S., "Coalition, cryptography, and stability: Mechanisms for coalition formation in task oriented domains, Alfred P. Sloan School of Management, Massachusetts Institute of Technology, (1994).
7. Fadaei, M., Tavakkoli-Moghaddam, R., Taleizadeh, A. and Mohammaditabar, D., "Cooperative benefit and cost games under fairness concerns", ***International Journal of Engineering, Transactions B: Applications***, Vol. 31, No. 8, (2018), 1250-1257. doi: 10.5829/ije.2018.31.08b.09.
8. Perozzi, B., Al-Rfou, R. and Skiena, S., "Deepwalk: Online learning of social representations", in Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. (2014), 701-710.
9. Taheri, A., Gimpel, K. and Berger-Wolf, T., "Sequence-to-sequence modeling for graph representation learning", ***Applied Network Science***, Vol. 4, (2019), 1-26. https://doi.org/10.1007/s41109-019-0174-8
10. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S. and Dean, J., "Distributed representations of words and phrases and their compositionality", ***Advances in Neural Information Processing Systems***, Vol. 26, No., (2013). doi: 10.48550/arXiv.1310.4546.
11. Salehi, S. and Pouyan, A., "Detecting overlapping communities in social networks using deep learning", ***International Journal of Engineering, Transactions C: Aspects***, Vol. 33, No. 3, (2020), 366-376. doi: 10.5829/ije.2020.33.03c.01.
12. Indrawan, G., Setiawan, H. and Gunadi, A., "Multi-class svm classification comparison for health service satisfaction survey data in bahasa", ***HighTech and Innovation Journal***, Vol. 3, No. 4, (2022), 425-442. doi: 10.28991/HIJ-2022-03-04-05.
13. Chakraborty, T., Dalmia, A., Mukherjee, A. and Ganguly, N., "Metrics for community analysis: A survey", ***ACM Computing Surveys (CSUR)***, Vol. 50, No. 4, (2017), 1-37. doi: 10.1145/3091106.
14. Cavallari, S., Zheng, V.W., Cai, H., Chang, K.C.-C. and Cambria, E., "Learning community embedding with community detection and node embedding on graphs", in Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. (2017), 377-386.
15. Van Laarhoven, T. and Marchiori, E., "Local network community detection with continuous optimization of conductance and weighted kernel k-means", ***The Journal of Machine Learning Research***, Vol. 17, No. 1, (2016), 5148-5175.
16. Liu, J., "Comparative analysis for k-means algorithms in network community detection", in Advances in Computation and Intelligence: 5th International Symposium, ISICA 2010, Wuhan, China, October 22-24, 2010. Proceedings 5, Springer. (2010), 158-169.
17. Ferreira, L.N., Pinto, A. and Zhao, L., "Qk-means: A clustering technique based on community detection and k-means for deployment of cluster head nodes", in The 2012 International Joint Conference on Neural Networks (IJCNN), IEEE. (2012), 1-7.
18. Niu, Y., Kong, D., Liu, L., Wen, R. and Xiao, J., "Overlapping community detection with adaptive density peaks clustering and iterative partition strategy", ***Expert Systems with Applications***, Vol. 213, (2023), 119213. https://doi.org/10.1016/j.eswa.2022.119213
19. Wu, Y., Fu, Y., Xu, J., Yin, H., Zhou, Q. and Liu, D., "Heterogeneous question answering community detection based on graph neural network", ***Information Sciences***, Vol. 621, (2023), 652-671. https://doi.org/10.1016/j.ins.2022.10.126
20. Chen, J., Li, Y., Yang, X., Zhao, S. and Zhang, Y., "Vghc: A variable granularity hierarchical clustering for community detection", ***Granular Computing***, Vol. 6, (2021), 37-46. https://doi.org/10.1007/s41066-019-00195-1
21. Yin, C., Zhu, S., Chen, H., Zhang, B. and David, B., "A method for community detection of complex networks based on hierarchical clustering", ***International Journal of Distributed Sensor Networks***, Vol. 11, No. 6, (2015), 849140. https://doi.org/10.1155/2015/849140
22. Zhou, L., Lü, K., Yang, P., Wang, L. and Kong, B., "An approach for overlapping and hierarchical community detection in social networks based on coalition formation game theory", ***Expert Systems with Applications***, Vol. 42, No. 24, (2015), 9634-9646. https://doi.org/10.1016/j.eswa.2015.07.023
23. Li, C., Bai, J., Wenjun, Z. and Xihao, Y., "Community detection using hierarchical clustering based on edge-weighted similarity in cloud environment", ***Information Processing & Management***, Vol. 56, No. 1, (2019), 91-109. https://doi.org/10.1016/j.ipm.2018.10.004
24. Newman, M.E., "Modularity and community structure in networks", ***Proceedings of the National Academy of Sciences***, Vol. 103, No. 23, (2006), 8577-8582. https://doi.org/10.1073/pnas.0601602103
25. Grover, A. and Leskovec, J., "Node2vec: Scalable feature learning for networks", in Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining. (2016), 855-864.
26. Wang, D., Cui, P. and Zhu, W., "Structural deep network embedding", in Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining. (2016), 1225-1234.
27. Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J. and Mei, Q., "Line: Large-scale information network embedding", in Proceedings of the 24th international conference on world wide web. (2015), 1067-1077.
28. Yang, C., Liu, Z., Zhao, D., Sun, M. and Chang, E.Y., "Network representation learning with rich text information", in IJCAI. Vol. 2015, (2015), 2111-2117.

29. Pan, S., Wu, J., Zhu, X., Zhang, C. and Wang, Y., "Tri-party deep network representation", ***Network***, Vol. 11, No. 9, (2016), 12.
30. Chang, J. and Blei, D., "Relational topic models for document networks", in Artificial intelligence and statistics, PMLR. (2009), 81-88.
31. McSweeney, P.J., Mehrotra, K. and Oh, J.C., "A game theoretic framework for community detection", in 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, IEEE. (2012), 227-234.
32. Zhou, L., Lü, K., Cheng, C. and Chen, H., "A game theory based approach for community detection in social networks", in Big Data: 29th British National Conference on Databases, BNCOD 2013, Oxford, UK, July 8-10, 2013. Proceedings 29, Springer. (2013), 268-281.
33. Hajibagheri, A., Alvari, H., Hamzeh, A. and Hashemi, S., "Social networks community detection using the shapley value", in The 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP 2012), IEEE. (2012), 222-227.
34. Avrachenkov, K.E., Kondratev, A.Y., Mazalov, V.V. and Rubanov, D.G., "Network partitioning algorithms as cooperative games", ***Computational Social Networks***, Vol. 5, No. 1, (2018), 1-28. doi: 10.1186/s40649-018-0059-5.
35. Zhou, X., Cheng, S. and Liu, Y., "A cooperative game theory-based algorithm for overlapping community detection", ***IEEE Access***, Vol. 8, (2020), 68417-68425. doi: 10.1109/ACCESS.2020.2985397.
36. Chen, W., Liu, Z., Sun, X. and Wang, Y., "A game-theoretic framework to identify overlapping communities in social networks", ***Data Mining and Knowledge Discovery***, Vol. 21, (2010), 224-240. https://doi.org/10.1007/s10618-010-0186-6
37. Narayanam, R. and Narahari, Y., "A game theory inspired, decentralized, local information based algorithm for community detection in social graphs", in Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), IEEE. (2012), 1072-1075.
38. Alvari, H., Hashemi, S. and Hamzeh, A., "Detecting overlapping communities in social networks by game theory and structural equivalence concept", in Artificial Intelligence and Computational Intelligence: Third International Conference, AICI 2011, Taiyuan, China, September 24-25, 2011, Proceedings, Part II 3, Springer. (2011), 620-630.
39. Havvaei, E. and Deo, N., "A game-theoretic approach for detection of overlapping communities in dynamic complex networks", ***International Journal of Computational Methods***, Vol. 1, (2016), 313-324. https: //doi.org/10.48550/arXiv.1603.00509
40. Zhao, X., Wu, Y., Yan, C. and Huang, Y., "An algorithm based on game theory for detecting overlapping communities in social networks", in 2016 International Conference on Advanced Cloud and Big Data (CBD), IEEE. (2016), 150-157.
41. Moscato, V., Picariello, A. and Sperli, G., "Community detection based on game theory", ***Engineering Applications of Artificial Intelligence***, Vol. 85, (2019), 773-782. https://doi.org/10.1016/j.engappai.2019.08.003
42. Wang, Y., Bu, Z., Yang, H., Li, H.-J. and Cao, J., "An effective and scalable overlapping community detection approach: Integrating social identity model and game theory", ***Applied Mathematics and Computation***, Vol. 390, (2021), 125601. https://doi.org/10.1016/j.amc.2020.125601
43. Zhou, L., Yang, P., Lü, K., Wang, L. and Chen, H., "A fast approach for detecting overlapping communities in social networks based on game theory", in Data Science: 30th British International Conference on Databases, BICOD 2015, Edinburgh, UK, July 6-8, 2015, Proceedings 30, Springer. (2015), 62-73.
44. Torkaman, A., Badie, K., Salajegheh, A., Bokaei, M.H. and Ardestani, S.F.F., "A four-stage algorithm for community detection based on label propagation and game theory in social networks", ***AI***, Vol. 4, No. 1, (2023), 255-269. https://doi.org/10.3390/ai4010011
45. Torkaman, A., Badie, K., Salajegheh, A., Bokaei, M.H. and Fatemi, S.F., "A hybrid deep network representation model for detecting researchers' communities", ***Journal of AI and Data Mining***, Vol. 10, No. 2, (2022), 233-243. https://doi.org/10.22044/jadm.2022.11243.2277
46. Myerson, R., "Game theory: Analysis of conflict harvard univ", ***Press, Cambridge***, Vol. 3, (1991). https://doi.org/10.2307/j.ctvjsf522
47. MacQueen, J., "Some methods for classification and analysis of multivariate observations", in Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Oakland, CA, USA. Vol. 1, (1967), 281-297.
48. Bholowalia, P. and Kumar, A., "Ebk-means: A clustering technique based on elbow method and k-means in wsn", ***International Journal of Computer Applications***, Vol. 105, No. 9, (2014). doi: 10.5120/18405-9674.
49. Newman, M.E. and Girvan, M., Mixing patterns and community structure in networks, in Statistical mechanics of complex networks. 2003, Springer.66-87.
50. Lusseau, D., "The emergent properties of a dolphin social network", ***Proceedings of the Royal Society of London. Series B: Biological Sciences***, Vol. 270, No. suppl_2, (2003), S186-S188. doi: 10.1098/rsbl.2003.0057.
51. Giles, C.L., Bollacker, K.D. and Lawrence, S., "Citeseer: An automatic citation indexing system", in Proceedings of the third ACM conference on Digital libraries. (1998), 89-98.
52. Ng, A., Jordan, M. and Weiss, Y., "On spectral clustering: Analysis and an algorithm", ***Advances in Neural Information Processing Systems***, Vol. 14, (2001).
53. Blondel, V.D., Guillaume, J.-L., Lambiotte, R. and Lefebvre, E., "Fast unfolding of communities in large networks", ***Journal of Statistical Mechanics: Theory and Experiment***, Vol. 2008, No. 10, (2008), P10008. doi: 10.1088/1742-5468/2008/10/P10008.
54. Pan, S., Hu, R., Long, G., Jiang, J., Yao, L. and Zhang, C., "Adversarially regularized graph autoencoder for graph embedding", arXiv preprint arXiv:1802.04407, (2018). https://doi.org/10.48550/arXiv.1802.04407
55. Wang, C., Pan, S., Hu, R., Long, G., Jiang, J. and Zhang, C., "Attributed graph clustering: A deep attentional embedding approach", arXiv preprint arXiv:1906.06532, (2019). https://doi.org/10.48550/arXiv.1802.04407
56. Liu, L., Kang, Z., Ruan, J. and He, X., "Multilayer graph contrastive clustering network", ***Information Sciences***, Vol. 613, (2022), 256-267. https://doi.org/10.1016/j.ins.2022.09.042
57. Qin, M., Jin, D., He, D., Gabrys, B. and Musial, K., "Adaptive community detection incorporating topology and content in social networks", in Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017. (2017), 675-682.
58. Kingma, D.P. and Ba, J., "Adam: A method for stochastic optimization", arXiv preprint arXiv:1412.6980, (2014). dhttps://doi.org/10.48550/arXiv.1412.6980
59. Yang, H., Kong, Q. and Mao, W., "A deep latent space model for graph representation learning", arXiv preprint arXiv:2106.11721, (2021). https://doi.org/10.48550/arXiv.2106.11721
60. Tatarkanov, A., Alexandrov, I., Muranov, A. and Lampezhev, A., "Development of a technique for the spectral description of curves of complex shape for problems of object classification", ***Emerging Science Journal***, Vol. 6, No. 6, (2022), 1455-1475. doi: 10.28991/ESJ-2022-06-06-015.

Persian Abstract

چکیده

امروزه با گسترش شبکه‌های پیچیده و حجم بالای داده‌های آن‌ها، یافتن روش‌های آشکارسازی دقیق هم‌پویه‌ها (جوامع) در این نوع شبکه‌ها از اهمیت بالایی برخوردارست. دشواری تجزیه و تحلیل این شبکه‌ها از حجم بالا و روابط پیچیده بین اعضای شبکه ناشی می شود. تحلیل روابط و سازوکار عمیق بین اعضا، تنها با در نظر گرفتن کلیت آن، دشوار است. روش‌های سنتی، برای تحلیل این شبکه‌ها مشکلات و محدودیت‌هایی نظیر استخراج ویژگی، وابستگی زیاد به تنظیمات اولیه، پیچیدگی محاسباتی بالا و نادیده گرفتن روابط و محتویات شبکه دارند. از نظر روابط و تعاملات بین افراد، محیط شبکه‌های پیچیده را می‌توان به بازی‌ای تشبیه کرد که در آن گره‌ها، به عنوان بازیکن یا عامل، تلاش می‌کنند تا بر اساس ویژگی‌های ساختاری یا معنایی مشابه به یک هم‌پویه بپیوندند یا از آن خارج شوند. در نتیجه، می‌توان برای آشکارسازی هم‌پویه‌ها از بازی‌های همکاری و غیر-همکاری بهره ببریم. اخیراً ادغام تکنیک‌های یادگیری ژرف و نظریه بازی‌ها جهت استخراج هم‌پویه‌ها بسیار مؤثر بوده است. تکنیک‌های یادگیری ژرف قابلیت‌های پیشرفته‌تری را در مهندسی ویژگی‌ها نشان داده‌اند و این فرآیند را خودکار نموده است. بنابریان، در این پژوهش، نویسندگان تلاش می‌کنند تا با کمک رویکردهای سنتی، یادگیری ژرف و همچنین بازی‌های همکاری و غیر-همکاری، هم‌پویه‌های منطقی و دقیقی را بر اساس ویژگی‌های ساختاری و محتوایی آشکارسازی کنند. کارایی این مطالعه با یافته‌های تجربی بر روی مجموعه داده‌های واقعی نشان داده شده است و تأیید می‌کند که روش پیشنهادی، به قدر کافی قادر به آشکارسازی هم‌پویه‌هایی است که معنی دارترند.