

Article

A Four-Stage Algorithm for Community Detection Based on Label Propagation and Game Theory in Social Networks

Atefeh Torkaman ¹, Kambiz Badie ², Afshin Salajegheh ^{1,*}, Mohammad Hadi Bokaei ³
and Seyed Farshad Fatemi Ardestani ⁴

¹ Department of Computer, South Tehran Branch, Islamic Azad University, Tehran 14778-93855, Iran

² E-Services and E-Content Research Group, IT Research Faculty, ICT Research Institute, Tehran 15916-34311, Iran

³ Department of Information Technology, ICT Research Institute, Tehran 15916-34311, Iran

⁴ Faculty of Management & Economics, Sharif University of Technology, Tehran 14588-89694, Iran

* Correspondence: a_salajegheh@azad.ac.ir

Abstract: Over the years, detecting stable communities in a complex network has been a major challenge in network science. The global and local structures help to detect communities from different perspectives. However, previous methods based on them suffer from high complexity and fall into local optimum, respectively. The Four-Stage Algorithm (FSA) is proposed to reduce these issues and to allocate nodes to stable communities. Balancing global and local information, as well as accuracy and time complexity, while ensuring the allocation of nodes to stable communities, are the fundamental goals of this research. The Four-Stage Algorithm (FSA) is described and demonstrated using four real-world data with ground truth and three real networks without ground truth. In addition, it is evaluated with the results of seven community detection methods: Three-stage algorithm (TS), Louvain, Infomap, Fastgreedy, Walktrap, Eigenvector, and Label propagation (LPA). Experimental results on seven real network data sets show the effectiveness of our proposed approach and confirm that it is sufficiently capable of identifying those communities that are more desirable. The experimental results confirm that the proposed method can detect more stable and assured communities. For future work, deep learning methods can also be used to extract semantic content features that are more beneficial to investigating networks.

Keywords: community detection; game theory; social networks; label propagation



Citation: Torkaman, A.; Badie, K.; Salajegheh, A.; Bokaei, M.H.; Ardestani, S.F.F. A Four-Stage Algorithm for Community Detection Based on Label Propagation and Game Theory in Social Networks. *AI* **2023**, *4*, 255–269. <https://doi.org/10.3390/ai4010011>

Academic Editors: Demos T. Tsahalos and Joaquín Torres-Sospedra

Received: 27 August 2022

Revised: 12 December 2022

Accepted: 12 December 2022

Published: 8 February 2023



Copyright: © 2023 by the authors. 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

Complex systems can be modeled as networks that are composed of subnets called communities, which are denser and more connected than the other parts of the network.

The structures of communities are directly related to the functions of a network. Identifying these structures is expected to yield helpful intuitions about the functional organization of the specific network. Community detection has been the most important topic in complex networks in recent decades, and one that has attracted many researchers, aiming to discover groups of nodes based on modular tendencies. The ability to detect communities provides deeper insight into the functionality of groups and how the networks are formed.

Different studies have various perspectives on detecting network communities, such as partitioning methods, modularity-based methods, factorization-based methods, etc. [1,2]. Some methods consider the global structure and the whole network's perspective, such as Infomap [3], Louvain [4], Fastgreedy [5], etc. while in many cases, they suffer from high cost and complexity.

On the other hand, some methods consider the local structure and extract the local information of the nodes [1–4]. These methods conquer the limitations of global structure-based methods. However, due to the significant development and growth of social networks, and the lack of access to global structure information, these types of algorithms may fall into a local minimum and may not have the desired accuracy, despite their low time complexity.

Therefore, creating a balance between global and local information, accuracy and time complexity is one of the crucial issues in community detection. Another important point is to achieve established and assured communities. In general, the environment of complex networks in terms of the relationships and interactions between their members can be considered as a game in which nodes as players or agents try to join or leave a community based on similar characteristics. On one hand, we can consider a cooperative environment as one where individuals with similar features interact with each other, they constitute the communities and attempt to promote the community's utility. On the other hand, it can also be considered a competitive space in which agents try to join or leave their communities and enhance their profits. A logical approach is necessary to interpret these relations. By borrowing game theoretical concepts from economics, this issue can easily be analyzed.

Game theory is a very useful mathematical means for studying strategic conditions and modeling the competition and cooperation between decision-makers to provide rational and optimal solutions in complicated situations. Generally, game theory is divided into two categories: cooperative and non-cooperative. A game that focuses on the member's cooperation is classified as a cooperative game [6], where every individual tries to improve the coalition's utility. Conversely, in non-cooperative games, individuals attempt to increase their own utilities and ignore the group's profits.

Hence, we came up with the idea to take into account both global and local information structures and cooperative and non-cooperative games to extract more satisfactory and assured communities. The proposed algorithm includes four stages: finding the important and central vertices, propagating the labels and identifying initial communities, merging these communities, and finally stabilizing them and assuring the nodes' allocation.

Taking this algorithm into account, the overall efficiency of the proposed algorithm increased, and the computational cost diminished remarkably. The subsequent parts of this paper are organized as follows: In Section 2, the literature about various approaches to community detection problems is reviewed. The basic concepts are mentioned in Section 3. The proposed model is brought up in Section 4. Analysis of the experimental results is discussed in Section 5, and finally concluding remarks and future works are presented in Section 6.

2. Related Work

A community is a subset of elements close to each other within their group rather than to the rest of the network. According to [1], approximately, the nodes of the same community exhibit similar characteristics, functions, and/or roles.

Community detection is one of the most fascinating research topics which has attracted the attention of many scientists in several fields, such as biology, statistics, economics, and computer science [1]. In general, community detection is an NP-Complete problem [2,3]. Various studies in the literature have tried extracting communities according to the global structure and whole network's perspectives, like Infomap [3], Louvain [4], Fastgreedy [5], etc., while in many cases, they suffer high cost and complexity.

Conversely, some methods consider the local structure and extract the local information of the nodes, and do not focus on the global knowledge of the network. Therefore, they are not as robust as the global algorithms. In addition, these types of methods may be caught into a local optimum. In contrast, they demonstrate less time complexity than global methods and are applicable to large-scale networks.

Some of the local algorithms are based on clustering, [5–18]. However, these approaches have some limitations, such as poor cluster descriptors and their high sensitivity to initial phase settings.

So, considering both global and local structures in community detection can be useful in eliminating the limitations of each method [19].

It is also noticeable that the communities detected based on the above-mentioned approaches may not be sufficiently qualified, to the extent that some nodes may be assigned to unreliable groups.

To circumvent this problem, game theoretical approaches to identifying communities have been proposed.

They have imagined the community detection issue as a game, in which each member rationally chooses a community and maximizes its score. Members of a community also attempt to enhance the group's utility.

Many approaches address the problem of community detection by using the non-cooperative game theory, and some others employ the cooperative one. Along the cooperative line, where individuals form a group based on the similarity of their communal interests, Mcsweeney et al. [20] considered each node as a player in a hedonic game, which tries to form fair and stable community structures. Zhou et al. [21], suggested the Shapley value to detect communities of a given social network. Additionally, they proposed a coalitional game for investigating communities based on the topological structure of nodes [16]. Hajibagheri and his colleagues [22] imagined each node as a rational individual trying to maximize the Shapley value. They considered community structure as Nash equilibrium. Two approaches from cooperative game theory based on the Myerson value and hedonic games were recommended in [23]. Both of them detected communities with different resolutions. Xu Zhou et al. [24] considered nodes as players who try to enhance the utility of their coalition by participating in a cooperative game. They proposed an edge weight computation for calculating the Shapley value of nodes and coalitions.

Regarding the non-cooperative aspect, according to Chen et al. [25], the utility of an agent is determined as gain and loss functions based on the modularity and community membership fee, respectively. Finally, the community structure was revealed by the local equilibrium of the game. Additionally, the authors in [26] regarded each vertex as an agent trying to join a community and assumed its utility as a linear function. Nash's stability guarantees the stability of communities. A framework based on the iterative game has been proposed in [27] for detecting communities in social networks. They considered nodes as rational agents who play the game to enhance their utilities. To reveal community structure, a weighted potential game was defined in [23]. Communities become stable as they reach the Nash equilibrium point. Zhao et al. [28] suggested Co-game, a game-theoretic approach for extracting community in real networks. This method produces finer-grained partitions in the detection process by combining individual games and equilibrium.

An algorithm, based on game theory for detecting communities in online social networks, was also proposed by Vincezo Moscato et al. [29]. They modeled the process of community formation as a game, in which each node as a player aims to maximize its goal. A new approach based on both cooperative and non-cooperative games for detecting communities was suggested by [30], which considered nodes as players in cooperative games who attempt to enhance the group's utility while engaging in a non-cooperative game to improve their utility.

In the first phase, this method, similarly to a hierarchical agglomerative method, considers a cooperative game in which individuals in a social network are modeled as rational players and aims to improve the utility of the group by cooperating with other players to form coalitions. In large datasets, this method, like other local and agglomerating approaches, typically suffers from high computational complexity.

The main problem with the existing approaches to game-theory-based community detection is that the game is initially started with single nodes. The main problem with the existing approaches to game-theory-based community detection is that the game is initially started with single nodes, with a large amount of comparisons between them, which in turn increases the computational cost, while in our approach the game is just considered for some extracted initial communities, thus leading to fewer comparisons between the nodes.

3. Basic Concepts

3.1. The Necessity of Representing the Network

Given a network $G = (V, E)$, $V = \{v_1, v_2, v_3, \dots, v_N\}$ is a set of nodes, where N is the number of nodes. $E = \{e_{ij}\}_{i,j=1}^N$ consists of the set of edges, where e_{ij} encodes the edge between v_i and v_j .

3.2. Community Detection

The community detection is to extract K communities, i.e., $C = \{C_1, C_2, \dots, C_k\}$, such that $K \ll N$ and $\cup_{k=1}^K C_k = V$. If these communities are non-empty, mutually exclusive subsets of V , i.e., $\forall i, j \in \{1, 2, \dots, K\}, i \neq j, C_i \cap C_j = \emptyset$, this is non-overlapping community detection, and nodes only can join one community. Conversely, this is entitled as overlapping communities, where nodes can join more than one community.

3.3. Sorensen Index

Sorensen–Dice coefficient, in short, the “Sorensen Index” [26], is a statistic that measures the similarity between two nodes by dividing the size of the intersection of their neighbor’s sets by the total number of their members (Equation (1)). The Sorensen index considers the degree of the two nodes and the number of their common neighbors.

The Sorensen index output is between 0 and 1:

$$S_{Sorensen}(u, v) = \frac{2 \times |N_u \cap N_v|}{d_u + d_v} \quad (1)$$

where N_u, N_v , and d_u, d_v are the neighbor sets and the degree of node i , respectively.

3.4. Game Theory Background

Game theory is a very useful mathematical means for studying strategic conditions between decision-makers to provide rational solutions in complicated situations. The environment of the relationships and interactions between the members of complex networks can be considered as a game in which nodes as players try to join or leave a community based on similar characteristics, where the decisions of one player influence the other player’s payoffs [25].

Let u_i be the utility function of node $i \in V$. For each community, $C_i \in C$, $u_i(C_i)$ demonstrates the utility of node i by being in the community C_i . Each node (player) tries to join a community and enhance its utility. It should be noted that the utility of any node depends on the community to which it belongs.

4. The Proposed Model

As mentioned before, the four-stage algorithm (FSA) has considered both global and local information structures with regard to the network. An overview of the proposed method can be seen in Figure 1.

Let us say, in the first step of the proposed algorithm, important nodes are determined according to their degree and relative distance. In the second step, the initial communities are detected based on the label propagation method. Next, the extracted communities are stabilized by the cooperative game, and finally, the non-cooperative game is applied to these clusters to ensure a rational allocation of nodes to the established communities.

So, identifying central nodes, label propagation, stabilizing the extracted communities, and ensuring the rational allocation of nodes to the established communities, can be considered the main characteristics of our approach.

Out of the above stages, the first two stages have been observed in previous works [25,27].

However, to be assured that the extracted communities would be stable, the third stage has been added based on the idea of a cooperative game. However, there are cases where a limited number of nodes may exist that could affiliate with a variety of different communities. In these cases, the fourth stage, which is based on the idea of a

non-cooperative game, would help us to ensure the rational allocation of these nodes to the deserving/properly established communities.

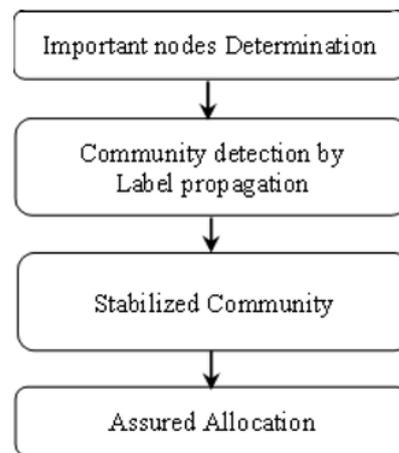


Figure 1. The steps of the proposed four-stage algorithm (FSA).

One of the main problems of community detection is to discover communities which are sufficiently qualified. Therefore, inventing an algorithm that can assign nodes to reliable groups is one of the most important topics in complex networks, such as social networks. Therefore, the proposed algorithm attempts to obtain high-quality and reliable communities by relying on cooperative and non-cooperative games.

4.1. Important Nodes Determination

Important nodes in a community have a high degree of surrounding neighbors. Approximately, the nodes with this characteristic are more likely to be the communities' centers. Speaking intuitively, the distances between important nodes are often far apart. Therefore, it can be assumed that the distance between two important nodes is not less than the average network distance. The average distance of graph G is [25]:

$$Avd = \frac{2}{n(n-2)} \sum_{u,v \in V} d(u,v) \quad (2)$$

where $d(u,v)$ is the shortest distance between u and v .

Then, rank all nodes by their degrees; $B = \{v_1, v_2, \dots, v_n\}$.

Then, let C be the set of important nodes. Initially, the highest-ranking node is settled in C :

$$C = \{v_1\}$$

$$\forall v_j \in B \ \& \ v_j \notin C, \ \forall v_i \in C, \ \text{if } d(v_i, v_j) \geq Avd, \ \text{then } C \cup \{v_j\}.$$

Repeat this until the distance between two nodes is not greater than the average distance of the graph (Equation (2)).

4.2. Community Detection by Label Propagation

Nodes in a community have similar characteristics and common interests and are more connected to each other than the rest of the network.

Having identified the important nodes in the network, the remaining nodes join the communities according to the Sorensen index, which is a useful index for comparing similarities between the samples [26].

Now, if we assume that every important node in C corresponds to an identified community, then we label every node $u \in V - C$ as the node $v \in C$ if u and v are neighbors:

$$S_{Sorenson}(u, v) = \max_{v_i \in C} \max_{v_j \in N(v_i)} S_{Sorenson}(v_i, v_j) \quad (3)$$

So, label u according to v . Repeat this process until all nodes have been labeled and assigned to the initial communities.

4.3. Stabilized Community

After the initial communities were formed by Label propagation, it is worth observing that some communities have only one member—in other words, they are sparse and do not have suitable quality—and these single nodes should join other communities with more nodes. For the sake of improving the quality, it is necessary to merge and reduce the initial communities. In this regard, we utilized the advantages of cooperative (coalitional form) games with a transferable utility, which is assumed that the earnings of a coalition (utility) can be distributed among the individuals in any conceivable way [28].

The reason for using cooperative game theory is that members of a community, based on their commitment towards the entire community, try to obtain a higher utility through cooperation. In other words, nodes in a network are modeled as logical agents (players), which try to form coalitions (communities) and cooperate to improve the group's utility.

Coalitions with single or fewer nodes join larger groups according to the utility measurement. The community, which gains the highest utility, is selected as the final community, and the single node, or the group with fewer members, joins it. Merging operations will continue until the utility no longer improved the utility of the merged coalitions. In this situation, the game has reached an equilibrium, and accordingly no coalition is willing to merge with the others. In this way, the number of communities will reduce, and high-quality coalitions will be obtained. In other words, the communities are stabilized.

Given S_i be a coalition of $G = \langle V, E \rangle$. The utility function $u(S_i)$ of S is:

$$u(S_i) = \sum_{S \in S} \left(\frac{e(S_i)}{|E|} - \left(\frac{D(S_i)}{2 \cdot |E|} \right)^2 \right) \quad (4)$$

where $|E|$ is the total number of edges in G , $e(S_i)$ is the number of edges that connect nodes within S_i , and $D(S_i)$ is the sum of the degrees of the nodes within S_i . $u(S_i)$ is based on Newman's modularity metric (Q) [25]. The modularity metric is one of the famous metrics that has been used in many kinds of research to measure the quality of the community structure in networks. The main idea of this index is based on comparing edge density within communities with the expected number in a random network. Thus, the value of 1 means that a network community structure has the highest possible strength [29].

Stable coalitions: A community is a stable coalition which is not eager to participate in the merged operation to improve its utility. In other words, S_i tries to join S_j if $u(S_i) > u(S_i + S_j) \forall S_j \neq S_i$ then it prefers to stay in the previous situation and with no further will to join S_j .

Utility increment: In the merge operation of S_j , let S_{ij} be a super-coalition of S_i obtained by the merged operation, so the utility increment of S_i is defined by $\Delta u(S_i, S_{ij}) = u(S_i, S_{ij}) - u(S_i)$. It means that the utility of S_i should increase within the merging coalition.

Generally, if $\Delta u(S_i, S_{ij}) > 0$ and $\Delta u(S_j, S_{ij}) > 0$ and then two communities are joined, the newly joined communities are added to a new list $Y = \{C_1, C_2, \dots, C_n\}$ which includes stable coalitions.

Stable coalitions are products of an equilibrium state for coalitions in which no group of agents has an interest in further merging operations.

It should be noted that coalitions with at least one edge in between are merged.

4.4. Assured Allocation

Having attained the set of stabilized communities, the non-cooperative game then takes place. The nodes may not be in their exact coalition. In this game, each node is considered a selfish agent, which attempts to join or leave a coalition from γ (stabilized communities' structure) based on its utility measurement. If by joining a coalition its utility increases, then it will leave the current coalition and join the new one.

Utility function of an agent: Let $x \in V$, $C_i \in \gamma$ the utility function is as [25]:

$$u_x(C_i) = \frac{e(x, C_i)}{d(x)} \quad (5)$$

$e(x, C_i)$ is the number of edges between x and coalition C_i . $d(x)$ is the degree of x . $u_x(C_i)$ measures the closeness between x and the targeted community C_i . The higher value of $u_x(C_i)$, the more similarity that exists between x and C_i .

Join and Leave: node x joins the community C_i

$$C_i \leftarrow C_i + \{x\}$$

If $x \notin C_i$ and $u_x(C_i) \geq \alpha$, x joins C_i .

Node x leaves its community C_n and joins community C_i :

$$C_n \leftarrow C_n - \{x\}$$

if $x \in C_n$ and $u_x(C_n) < \beta$.

α and β are the lower and upper bounds of the utility value of x , respectively.

The Four-Stage Algorithm (FSA) is described in Table 1.

As mentioned before, after identifying the important nodes in the network, the remaining nodes join the initial communities according to the Sorensen similarity index.

The cooperative game initiates between these communities. Given S_i, S_j two communities in S , if the utility of joining these two communities (S_{ij}) is greater than the utility of each of them, then the joining operation occurs and S_{ij} is added to a new list Υ , and the two communities are merged. Once equilibrium has been achieved, the utility of each group cannot be demonstrated. The stable communities are then identified; however, some nodes may not satisfy their utilities, and they begin to play the non-cooperative game to improve their utilities.

Each node evaluates the other communities and calculates its utility if it joins them. If the value is more than ω and lower than ε , the node leaves its current position and joins the new community. The algorithm ends when the agents are not interested in joining other communities to improve their utility values and are interested in staying in their current situation.

Since the cooperative game runs on the results of the initial clustering rather than singleton nodes, the complexity is reduced. In addition, because the non-cooperative game applied on the stabilized clusters has been achieved by the cooperative game, the nodes are most likely to be in their exact coalitions and therefore there would have been no intention to change membership in their community due to improving their utilities.

Table 1. The pseudo-code for Four-Stage Algorithm (FSA).

Important nodes Determination
1:Input: An undirected and unweighted network $G = (V, E)$ 2: Output: The important node's set $C = \{v_1, v_2, \dots, v_n\}$ 3: $C = \{v_1\}$ 4:for all $(v_j \in V, v_j \notin C)$ do 5: if $(d(v_j, v_1) \geq Avd)$ 6: $C = C \cup \{v_j\}$ 7: end if 8:Return C 9:end for
Community Detection (Label Propagation)
1:Input: Ranking nodes $C = \{v_1, v_2, \dots, v_n\}$ 2: Output: The communities $S = \{S_1, S_2, \dots, S_n\}$ 3: $S_i = \{v_{i_j}\}, v_{i_j} \in C$ 4:for all $(u \in S$ and $v \in V - C)$ do 5: if $(S_{Sorenson}(u, v) = \text{true})$ 6: $S_i = S_i \cup \{v\}$ 7: end if 8: $S = S \cup S_i$ 9: Return S 10:end for
Community Combination (Cooperative game)
1:Input: The initial communities $S = \{S_1, S_2, \dots, S_n\}$ 2:Output: Reduced and stabilized communities $\gamma = \{C_1, C_2, \dots, C_n\}$ 3: $\gamma = \{\}$ 4: for all $(S_i, S_j \in S$ and $S_i \neq S_j)$ do 5: if $\Delta u(S_{ij}) > \Delta u(S_j) \& \Delta u(S_j) > 0$ then 6: $\gamma = \{S_{ij}\} - \{S_i\} - \{S_j\}$ 7: else 8: Return γ 9: end else 10: end if 11: end for (Repeat until no coalition willing to join the other one to improve itself)
Assured Allocation (non-Cooperative game)
1:Input: The reduced and stabilized communities achieved by cooperative game $\gamma = \{C_1, C_2, \dots, C_n\}$ 2:Output: Assured node allocation and final stable community structure $C = \{C_1, C_2, \dots, C_n\}$ 3: $\delta = \{\}$ 4:for all $(x \in C_i)$ do 5: $\delta = C - C_i$ 6: for all $(C_j \in \delta)$ do 7: if $(\Delta u_x(C_i)) > \omega$ 8: $C_j = C_j + \{x\}$ 9: end if 10: if $(\Delta u_x(C_j)) < \epsilon$ 11: $C_i = C_i - \{x\}$ 12: end if 13: Return C_i, C_j 14:end for (Repeat until nodes do not eager to join new community and leave their current communities)

5. Analysis of the Experimental Results

To evaluate the capabilities and effectiveness of the proposed approach, the experiments are conducted on real networks with/without the ground truth and the benchmark network of Lancichinetti and Fortunato [9]. The outcomes of the four-stage algorithm are compared by seven other community detection methods: Three-stage algorithm (TS) [25], Louvain [4], Infomap [3], Fastgreedy [5], Walktrap [30], Eigenvector [31], and Label propagation (LPA) [32].

Before debating on the experimental results, two famous functions for evaluating the proposed algorithm are introduced as follows:

Normalized Mutual Information (NMI): This is a well-known approach for evaluating the performance of community detection algorithms, which determines the amount of similarity between the partition proposed by the algorithm and the desired partition [5]. The NMI between two identical partitions is 1 [33].

The standard normalized mutual information (NMI) metric defined in [33], is determined as follows:

$$I_{norm}(X, Y) = \frac{2I(X, Y)}{H(X) + H(Y)} \quad (6)$$

where $I(X, Y)$ is the mutual information between X and Y . $H(X)$ and $H(Y)$ are the entropy of X , Y . If the communities of X and Y are independent, then knowing X does not provide information about Y , therefore $NMI(X, Y) = 0$.

Modularity: This is a famous evaluation index, proposed by Newman and Girvan [34], for measuring the quality of the community structure in networks. The main idea of this index is based on comparing edge density within communities with the expected number in a random network.

The definition of modularity is:

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - \frac{d(v_i)d(v_j)}{2m}) \delta(C_i, C_j) \quad (7)$$

where m is the total number of edges, A_{ij} is the adjacency matrix, $d(v_i)$ is the degree of the v_i , and δ is an indicator function which is 1 if i and j are in the same community ($C_i = C_j$) and output 0 if they are in different communities.

The modularity value ranges between 0 and 1. If the whole graph is assumed as a community, the modularity value would be zero. A higher value of Q indicates a better community structure.

5.1. Real Networks with Ground Truth

In this research, the following four real networks with ground truth are used to test the efficiency and accuracy of the proposed algorithm.

Dolphin Network [34]: This network includes 62 nodes and 159 edges, which represent the relationships between two groups of dolphins.

Zachary Karate Club Network [4]: It consists of 34 nodes and 79 edges that were set between the individuals who intend to join one of the two clubs.

American College Football Network [35]: This network originates from the United States college football. It consists of 115 nodes and 616 edges. The team represented as nodes and edges have defined the regular season games between two related teams.

Polbooks network [5]: It includes 105 nodes and 882 links. The network consists of the US political books' data which were recorded in 2005 by Adamic and Glance.

Table 2 shows the NMI and the modularity values of only Label propagation, after applying the cooperative game and finally after running the non-cooperative game in the four real networks with ground truth.

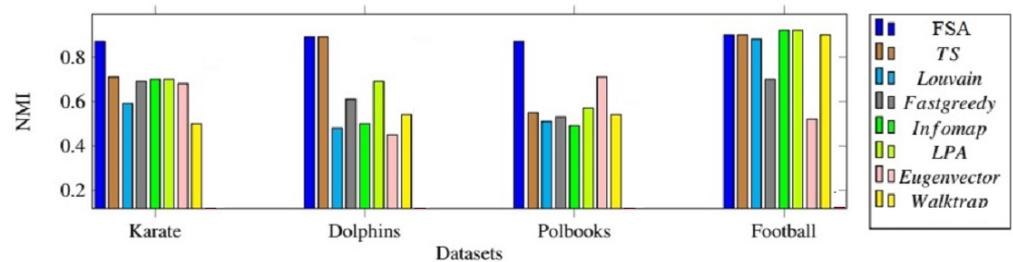
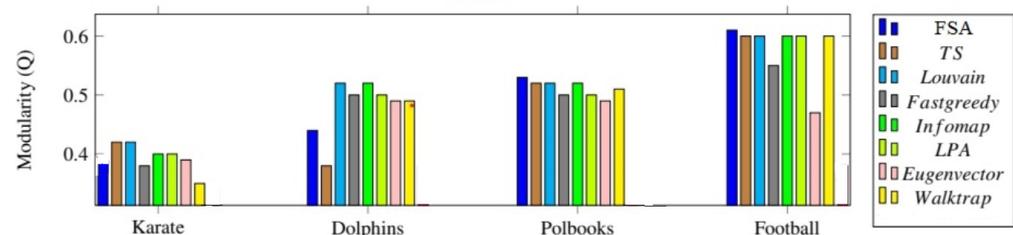
Table 2. The performance for each step of the FSA algorithm in the networks with ground truth.

Dataset	NMI			Modularity		
	Label Propagation	Cooperative Game	Non-Cooperative Game	Label Propagation	Cooperative Game	Non-Cooperative Game
Karate	0.3428	0.6948	0.8737	0.0015	0.2797	0.2890
Dolphin	0.2672	0.4888	0.8649	0.0119	0.2864	0.2991
Polbooks	0.3363	0.5036	0.8701	0.0064	0.0785	0.0884
Football	0.6845	0.5003	0.7261	0.0058	0.3705	0.3924

Table 2 indicates that the Label propagation method does not work properly in this situation. Because some obtained communities have low quality and are sparse, they need to merge with the other strong ones to qualify for the final communities. After applying the cooperative game to the Label propagation results, the NMI value and the modularity were improved due to the cluster merging process and stabilization point achievement.

As we may see, after running the non-cooperative game on the results of the cooperative approach, promising results were obtained. Once the equilibrium point is achieved, the condition of all nodes and communities stabilizes.

Figures 2 and 3 show the modularity and NMI values for the applied algorithms in the four real networks, respectively.

**Figure 2.** The NMI results for the Four-Stage Algorithm (FSA) and other approaches in the networks with ground truth.**Figure 3.** Modularity results for the Four-Stage Algorithm (FSA) and other approaches in the networks with ground truth.

As can be seen, the Four-Stage Algorithm (FSA) yielded better results in terms of NMI, particularly for the karate and polbooks networks. In addition, for the other datasets, the four-stage algorithm remains competitive.

In terms of modularity, the Four-Stage Algorithm (FSA) is better than other methods in dolphin, Football, and polbooks networks.

According to the results in Table 3, in most cases, the FSA method works much better than other clustering methods in the NMI and modularity. TS algorithm is in second place in this list, which has extracted four communities for the polbooks dataset, while the number of communities is two in the ground truth.

Table 3. Shows that the FSA algorithm has detected a close number of communities to the ground truth.

Methods	Networks				
	Evaluation Approaches	Karate	Dolphin	Polbooks	Football
Ground Truth	Q	0.37	0.38	0.41	0.55
	C	2	2	3	12
FSA	Q	0.37	0.44	0.53	0.61
	NMI	0.87	0.89	87	0.9
TS	C	2	2	4	10
	Q	0.42	0.38	0.52	0.6
Louvain	NMI	0.71	0.89	0.55	0.9
	C	4	2	4	10
Fast Greedy	Q	0.42	0.52	0.52	0.6
	NMI	0.59	0.48	0.51	88
Infomap	Q	0.38	0.5	0.5	0.55
	NMI	0.69	0.61	0.53	0.7
LPA	Q	0.4	0.52	0.52	0.6
	NMI	0.7	0.5	0.49	0.92
Eigenvector	Q	0.4	0.5	0.5	0.6
	NMI	0.7	0.69	0.57	0.92
Walktrap	Q	0.39	0.49	0.49	0.47
	NMI	0.68	0.45	0.71	0.52
	Q	0.35	0.49	0.51	0.6
	NMI	0.5	0.54	0.54	0.9

5.2. Real Networks without Ground Truth

Additionally, to evaluate the efficiency and accuracy of the four-stage and seven other algorithms, three real networks without ground truth are investigated as follows:

Lesmis network [36]: This undirected network contains co-occurrences of characters in Victor Hugo's novel "Les Misérables", as compiled by Donald Ervin Knuth. Nodes represent characters and the edge between two nodes shows that these two characters appeared in the same chapter of the book.

Adjnoun network [31]: A network of common adjective and noun adjacencies for the novel "David Copperfield" by Charles Dickens, as described by M. Newman. Nodes represent the most common adjectives and nouns in the novel. Edges connect each pair of words that are in adjacent positions in the text of the book.

Jazz network [37]: is the collaboration network between jazz musicians. The nodes are jazz musicians and the edges indicate the cooperation of two musicians in a band.

Table 4 represents the results of the modularity (Q) [38] and the number of detected communities (C) in real datasets without ground truth. In the Lesmis dataset, the maximum modularity, $Q = 0.56$, belongs to the Louvain algorithm, while in the Adjnoun and Jazz datasets, the proposed method has a more promising result.

Table 4. The performance and the number of extracted communities in real networks without ground truth.

Networks	FSA		TS		Louvain		FastGreedy		Infomap		LPA		Eigenvector		Walktrap	
	C	Q	C	Q	C	Q	C	Q	C	Q	C	Q	C	Q	C	Q
Lesmis	6	0.55	6	0.54	6	0.56	5	0.50	9	0.55	8	0.53	6	0.55	8	0.52
Adjnoun	7	0.31	7	0.29	7	0.29	7	0.29	2	0.01	10	0.24	1	0.00	25	0.22
Jazz	3	0.45	3	0.44	4	0.44	4	0.44	7	0.28	3	0.39	2	0.28	11	0.44

The number of extracted communities in Table 4 reveals that in the Lesmis dataset, FSA, TS, Louvain, and eigenvector achieve good results. FSA, TS, Louvain, and Fastgreedy detected an identical number of communities in the Adjnoun network, and in the Jazz network, the FSA algorithm detected three communities that are close to TS, Louvain, and Fastgreedy. However, its larity has been slightly improved in two recent datasets. Walktrap has performed very differently in these networks than the other algorithms.

5.3. Time Analysis of the Proposed Algorithm

The calculated running time (in seconds) for the FSA algorithm and other algorithms in real-world datasets is shown in Table 5. All of the experiments are implemented on a desktop PC with an Intel Core i7 CPU (3.4 GHz) and 8 GB RAM.

Table 5. Comparison of the runtime (in seconds) of the proposed algorithm in real datasets.

Networks	FSA	Louvain	Infomap	LPA	Walktrap
Karate	0.0007	0.1061	0.0200	0.0041	0.0039
Football	0.0009	0.1082	0.0170	0.0032	0.0059
Dolphin	0.0010	0.1078	0.0181	0.0023	0.0041
Polbooks	0.0014	0.1101	0.0208	0.0031	0.0049
Lesmis	0.0011	0.1090	0.0095	0.0029	0.0051
Jazz	0.0034	0.1890	0.1098	0.0078	0.0090
Adjnoun	0.0021	0.1001	0.0971	0.0034	0.0058

5.4. Benchmark Networks

In this section, a series of benchmark networks are applied according to the method of Lancichinetti, Fortunato, and Raddichi (LFR) [9]. These networks have power law distributions and to some extent are suitable for evaluating the performance of community detection algorithms [39–45].

The parameters used in this research are the same as [9] except the number of nodes in the network (n) which is considered as $n = 50, 100, 150, 200$.

The power law exponent for the size of communities: $\beta = 1$.

As seen in Table 6, the NMI value for the label propagation method is lower than the other steps.

Table 6. The performance for each step of the FSA algorithm in the benchmark networks.

n	NMI			Modularity		
	Label Propagation	Coalition	Individual	Label Propagation	Coalition	Individual
50	0.3224	0.9049	0.9321	0.0244	0.5008	0.6127
100	0.4029	0.9267	0.9526	0.0010	0.5320	0.5340
150	0.4849	0.9602	0.9731	0.0168	0.6972	0.7321
200	0.5413	0.8387	0.9606	0.0299	0.6896	0.7487

In the cooperative game deployment step, the NMI value and the modularity were improved due to the merging process and reaching the stabilizing communities. Finally, using the non-cooperative game step, promising values of NMI and modularity were obtained, as each node tries to improve its utility and stabilize its position in the communities.

According to Figure 4a, initially, with the increase in ε , the probability of removing nodes from the clusters increases, and those nodes which have fewer connections with each cluster are removed. As a result, the NMI value will increase.

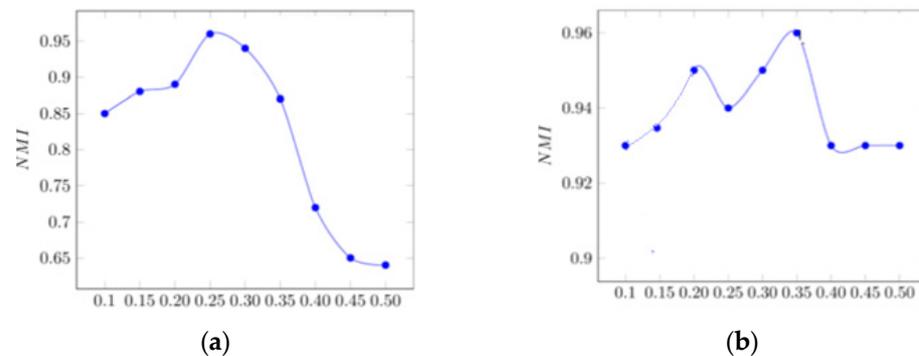


Figure 4. The NMI values of the four-stage algorithm on the benchmark networks based on (a) ε , (b) ω .

However, after a while, when the ε value exceeds the threshold (approximately 0.24), nodes that have more intra-cluster communications will also be removed from that cluster and the NMI value will subsequently decrease sharply.

In Figure 4b, the optimal value of ω is approximately 0.35, and for the higher values, the NMI value will practically not change, and nodes will rarely join the clusters.

6. Concluding Remarks and Future Works

In this paper, we suggested a powerful and effective community detection approach that incorporates global and local information into community detection. First, important nodes in the networks are determined and then label propagation is applied to find the initial clusters. However, some sparse clusters require merging with other strong types and their situation stabilizes; therefore, a cooperative game is used. Ultimately, this guarantees a rational allocation of nodes to established communities, where a non-cooperative game is performed on each node.

The experimental results confirm the performance of our approach and demonstrate that cooperative and non-cooperative game approaches boost each other to detect more stables and assured communities.

We evaluated the proposed method on several standard real networks and benchmark datasets and compared the performance of the FSA method with other algorithms. There are several advantages of the proposed method, including the formation of high-quality communities, solving the problem of community quality dependency on the inappropriately selected node, and allocate nodes to their stable communities.

The proposed method can respond successfully to a wide range of real-world issues like real business recommendations, precision marketing, etc. The application of the FSA algorithm in healthcare is settled by its fundamental idea; it can be applied to the diagnosis and treatment of diseases in general, and mental diseases in particular. It can be used to identify the patients, enhance their conditions, and detect the contagion of the disease. By using this method, it is possible to easily solve the challenges of the supply and demand side, which improves the condition of patients.

Additionally, it is also suitable for recommending and finding friends on social networks.

The proposed algorithm has high performance in small- and medium-size data sets. Although it has limitations in large datasets, where the number of extracted communities is large, it takes time to detect suitable communities once we use game theory, especially when it comes to considering non-cooperative games.

Furthermore, it can only detect non-overlapping communities from unweighted and undirected networks.

For future work, some solutions, such as deep learning methods, can be used to extract features that are more beneficial to large networks, especially those with semantic content. Furthermore, overlapping communities and weighted networks would be considered.

Author Contributions: Conceptualization, A.T. and K.B.; methodology, A.T. and M.H.B. and S.F.F.A.; software, A.T.; validation, A.T.; formal analysis, A.T., K.B., M.H.B., A.S. and S.F.F.A.; investigation, A.T.; resources, A.T. and M.H.B.; data curation, A.T.; writing—original draft preparation, A.T. and K.B.; writing—review and editing, A.T. and K.B.; visualization, A.T., K.B. and A.S.; supervision, K.B., M.H.B. and A.S.; project administration, K.B., M.H.B. and A.S.; funding acquisition, A.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Coscia, M.; Giannotti, F.; Pedreschi, D. A classification for community discovery methods in complex networks. *Stat. Anal. Data Mining ASA Data Sci. J.* **2011**, *4*, 512–546. [[CrossRef](#)]
2. Fortunato, S. Community detection in graphs. *Phys. Rep.* **2009**, *486*, 75–174. [[CrossRef](#)]
3. Rosvall, M.; Bergstrom, C.T. Maps of information flow reveal community structure in complex networks. *arXiv* **2007**, arXiv:0707.0609.
4. Blondel, V.D.; Guillaume, J.-L.; Lambiotte, R.; Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, *2008*, P10008. [[CrossRef](#)]
5. Clauset, A.; Newman, M.E.J.; Moore, C. Finding community structure in very large networks. *Phys. Rev. E* **2004**, *70*, 066111. [[CrossRef](#)]
6. Guo, K.; He, L.; Chen, Y.; Guo, W.; Zheng, J. A local community detection algorithm based on internal force between nodes. *Appl. Intell.* **2019**, *50*, 328–340. [[CrossRef](#)]
7. Li, H.-J.; Bu, Z.; Li, A.; Liu, Z.; Shi, Y. Fast and Accurate Mining the Community Structure: Integrating Center Locating and Membership Optimization. *IEEE Trans. Knowl. Data Eng.* **2016**, *28*, 2349–2362. [[CrossRef](#)]
8. Ding, X.; Zhang, J.; Yang, J. A robust two-stage algorithm for local community detection. *Knowledge-Based Syst.* **2018**, *152*, 188–199. [[CrossRef](#)]
9. Whang, J.J.; Gleich, D.F.; Dhillon, I.S. Overlapping Community Detection Using Neighborhood-Inflated Seed Expansion. *IEEE Trans. Knowl. Data Eng.* **2016**, *28*, 1272–1284. [[CrossRef](#)]
10. Nash, J. Non-cooperative games. *Ann. Math.* **1951**, *54*, 286–295. [[CrossRef](#)]
11. Cavallari, S.; Zheng, V.W.; Cai, H.; Chang, K.C.-C.; 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, Singapore, 6–10 November 2017; pp. 377–386.
12. Chakraborty, T.; Dalmia, A.; Mukherjee, A.; Ganguly, N. Metrics for community analysis: A survey. *ACM Comput. Surv. (CSUR)* **2017**, *50*, 1–37. [[CrossRef](#)]
13. Liu, J. Comparative analysis for k-means algorithms in network community detection. In Proceedings of the International Symposium on Intelligence Computation and Applications, Wuhan, China, 22–24 October 2010; Springer: Berlin/Heidelberg, Germany, 2010; pp. 158–169. [[CrossRef](#)]
14. Ferreira, L.N.; Pinto, A.R.; Zhao, L. QK-means: A clustering technique based on community detection and K-means for deployment of cluster head nodes. In Proceedings of the 2012 International Joint Conference on Neural Networks (IJCNN), Brisbane, QLD, Australia, 10–15 June 2012; pp. 1–7. [[CrossRef](#)]
15. Van Laarhoven, T.; Marchiori, E. Local network community detection with continuous optimization of conductance and weighted kernel k-means. *J. Mach. Learn. Res.* **2016**, *17*, 5148–5175.
16. Lancichinetti, A.; Fortunato, S. Consensus clustering in complex networks. *Sci. Rep.* **2012**, *2*, 336. [[CrossRef](#)] [[PubMed](#)]
17. Zhang, S.; Wang, R.-S.; Zhang, X.-S. Identification of overlapping community structure in complex networks using fuzzy c-means clustering. *Phys. A Stat. Mech. Its Appl.* **2007**, *374*, 483–490. [[CrossRef](#)]
18. Chen, J.; Li, Y.; Yang, X.; Zhao, S.; Zhang, Y. VGHC: A variable granularity hierarchical clustering for community detection. *Granul. Comput.* **2019**, *6*, 37–46. [[CrossRef](#)]
19. Newman, M.E. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8577–8582. [[CrossRef](#)]

20. McSweeney, P.J.; Mehrotra, K.; Oh, J.C. A game theoretic framework for community detection. In Proceedings of the 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Istanbul, Turkey, 26–29 August 2012; pp. 227–234.
21. Zhou, L.; Lü, K.; Cheng, C.; Chen, H. A game theory based approach for community detection in social networks. In Proceedings of the British National Conference on Databases, Oxford, UK, 8–10 July 2013; Springer: Berlin/Heidelberg, Germany, 2013; pp. 268–281.
22. Hajibagheri, A.; Alvari, H.; Hamzeh, A.; Hashemi, S. Social networks community detection using the shapley value. In Proceedings of the 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP 2012), Shiraz, Iran, 2–3 May 2012; pp. 222–227.
23. Avrachenkov, K.E.; Kondratev, A.Y.; Mazalov, V.; Rubanov, D.G. Network partitioning algorithms as cooperative games. *Comput. Soc. Netw.* **2018**, *5*, 1–28. [[CrossRef](#)]
24. Zhou, X.; Cheng, S.; Liu, Y. A Cooperative Game Theory-Based Algorithm for Overlapping Community Detection. *IEEE Access* **2020**, *8*, 68417–68425. [[CrossRef](#)]
25. Alvari, H.; Hashemi, S.; Hamzeh, A. Detecting overlapping communities in social networks by game theory and structural equivalence concept. In Proceedings of the International Conference on Artificial Intelligence and Computational Intelligence, Taiyuan, China, 24–25 September 2011; Springer: Berlin/Heidelberg, Germany, 2011; pp. 620–630.
26. Narayanam, R.; 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), Tsukuba, Japan, 11–15 November 2012; pp. 1072–1075.
27. Havvaei, E.; Deo, N. A game-theoretic approach for detection of overlapping communities in dynamic complex networks. *Int. J. Math. Comput. Methods* **2016**, *1*, 313–324.
28. Zhao, X.; Wu, Y.; Yan, C.; Huang, Y. An algorithm based on game theory for detecting overlapping communities in social networks. In Proceedings of the 2016 International Conference on Advanced Cloud and Big Data (CBD), Chengdu, China, 13–16 August 2016; pp. 150–157.
29. Moscato, V.; Picariello, A.; Sperli, G. Community detection based on game theory. *Eng. Appl. Artif. Intell.* **2019**, *85*, 773–782. [[CrossRef](#)]
30. Zhou, L.; Yang, P.; Lü, K.; Wang, L.; Chen, H. A fast approach for detecting overlapping communities in social networks based on game theory. In Proceedings of the British International Conference on Databases, Oxford, UK, 10–12 July 2015; Springer: Berlin/Heidelberg, Germany, 2015; pp. 62–73.
31. Sorensen, T.A. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. *Biol. Skar.* **1948**, *5*, 1–34.
32. Myerson, R.B. *Game Theory: Analysis of Conflict*; Harvard University Press: Cambridge, MA, USA, 1997.
33. You, X.; Ma, Y.; Liu, Z. A three-stage algorithm on community detection in social networks. *Knowl.-Based Syst.* **2019**, *187*, 104822. [[CrossRef](#)]
34. Newman, M.E.; Girvan, M. Mixing patterns and community structure in networks. In *Statistical Mechanics of Complex Networks*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 66–87.
35. Pons, P.; Latapy, M. Computing communities in large networks using random walks. In Proceedings of the International Symposium on Computer and Information Sciences, Istanbul, Turkey, 26–28 October 2005; Springer: Berlin/Heidelberg, Germany, 2005; pp. 284–293.
36. Newman, M.E.J. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E* **2006**, *74*, 036104. [[CrossRef](#)] [[PubMed](#)]
37. Raghavan, U.N.; Albert, R.; Kumara, S. Near linear time algorithm to detect community structures in large-scale networks. *Phys. Rev. E* **2007**, *76*, 036106. [[CrossRef](#)]
38. Newman, M.E.J.; Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **2004**, *69*, 026113. [[CrossRef](#)]
39. Danon, L.; Diaz-Guilera, A.; Duch, J.; Arenas, A. Comparing community structure identification. *J. Stat. Mech. Theory Exp.* **2005**, *2005*, P09008. [[CrossRef](#)]
40. Lusseau, D. The emergent properties of a dolphin social network. *Proc. R. Soc. B Boil. Sci.* **2003**, *270*, S186–S188. [[CrossRef](#)]
41. Zachary, W.W. An Information Flow Model for Conflict and Fission in Small Groups. *Anthropol. Res.* **1977**, *33*, 452–473. [[CrossRef](#)]
42. Lancichinetti, A.; Fortunato, S.; Radicchi, F. Benchmark graphs for testing community detection algorithms. *Phys. Rev. E* **2008**, *78*, 046110. [[CrossRef](#)]
43. Chen, M.; Kuzmin, K.; Szymanski, B.K. Community Detection via Maximization of Modularity and Its Variants. *IEEE Trans. Comput. Soc. Syst.* **2014**, *1*, 46–65. [[CrossRef](#)]
44. Aghaalizadeh, S.; Afshord, S.T.; Bouyer, A.; Anari, B. A three-stage algorithm for local community detection based on the high node importance ranking in social networks. *Phys. A Stat. Mech. Its Appl.* **2020**, *563*, 125420. [[CrossRef](#)]
45. Peters, H. *Game Theory: A Multi-Leveled Approach*; Springer: Berlin/Heidelberg, Germany, 2015.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.