Computing and Informatics, Vol. 44, 2025, 43–70, doi: 10.31577/cai_2025_1_43

# ADAPTIVE NON-OVERLAPPING COMMUNITY DETECTION BASED ON GRAVITATIONAL FIELD STABILITY IN SOCIAL NETWORKS

Meizi Li, Yanmei Gu

*College of Information, Mechanical and Electrical Engineering*
*Shanghai Normal University, Shanghai, 200234, China*
*&*
*Shanghai Engineering Research Center of Intelligent Education and Bigdata*
*Shanghai Normal University, Shanghai, 200234, China*
*&*
*The Research Base of Online Education for Shanghai Middle and Primary Schools*
*Shanghai, 200234, China*

Qianqian Zhai

*College of Information, Mechanical and Electrical Engineering*
*Shanghai Normal University, Shanghai, 200234, China*
*&*
*Shanghai Engineering Research Center of Intelligent Education and Bigdata*
*Shanghai Normal University, Shanghai, 200234, China*

Xiaoyang Guo

*Shanghai Newtouch Software Co., Ltd., Shanghai, China*

Zhonghua Zheng

*Anhui Boryou Information Technology Co., Ltd., Shanghai, China*

Chang Guo*

*College of Information, Mechanical and Electrical Engineering*
*Shanghai Normal University, Shanghai, 200234, China*
*&*
*Shanghai Engineering Research Center of Intelligent Education and Bigdata*
*Shanghai Normal University, Shanghai, 200234, China*
*&*
*The Research Base of Online Education for Shanghai Middle and Primary Schools*
*Shanghai, 200234, China*
*e-mail:* guochang@shnu.edu.cn

**Abstract.** Community structure is a common feature of social networks and many community discovery algorithms have emerged through the study of this feature. The gravitational field model is an effective method to realize community division. However, the current gravitational field model lacks a comprehensive consideration of field properties such as the internal stability of the gravitational field. Therefore, in this paper, we define and quantify the attributes of the gravitational field by taking advantage of the field's strength in describing the joint action of groups. Then, we propose a social network gravitational field community detection model (GF-CDM). GF-CDM selects the field kernel node based on a random walk and then presents an adaptive expansion function of fusion field stability to divide the observable network into overlapping and non-overlapping clusters. The model was evaluated on four real network datasets and five artificial network datasets of different sizes. Experimental results show that our proposed model outperforms the other four benchmark algorithms in modularity, ARI index, and field average stability, which can improve the quality of cluster division.

**Keywords:** Community structure, social network, internal stability, gravitational field model, cluster division

## 1 INTRODUCTION

In social networks, individuals are grouped around a common interest, this grouping is called "community" [1]. It is significant to discover various communities in social networks, and potential relationships can be mined from them. With the continuous deepening of research on social networks, community structure, as an important attribute in complex networks, has been increasingly noticed. The community struc-

* Corresponding author

ture defines that nodes in the same community are closely connected, while nodes in different communities are sparsely connected [2]. The concept of a cluster is similar to a community, both describing a collection of nodes in a social network. A cluster is a group of nodes gathered together based on similarities or inter-node interactions, focusing on functional or algorithmic grouping of nodes based on specific criteria. At present, academic research has proposed several community detection methods for the characteristics of complex network clusters, which can be generally classified into global community detection and local community detection methods. Global community detection uses the global information of the entire social network to divide communities, which is difficult to apply in larger networks [3]. Local community detection uses local information to find the community where the seed node is located [4]. Local methods improve the efficiency of the algorithm using widely in the large network with dynamic changes and have a broader space for development [5]. The community partition method based on seed nodes for local expansion is to optimally partition the seed nodes and their neighbouring nodes into correct communities [6]. The fitness function describes the tightness of a group of node connections. A cluster is composed of a group of nodes that can obtain the maximum fitness function value.

Extended cluster partitioning method based on fitness function selected the node in which fitness value is maximum and positive to join the cluster. The nodes with negative fitness values in the updated community were removed. These two processes are iterated until all the neighbors of the community have negative fitness function values. This method improves the quality of the detected communities. Due to the increasing amount of data, the Global community detection is difficult to use for larger networks. Therefore, it is particularly important to improve the accuracy of cluster division with limited local information.

At present, cluster division is one of the popular projects in the research of gravitational field models in social networks. The gravitational field model in physics regards each object in space as a particle with a certain mass. Similarly, there is a virtual gravitational field in the social network that any object in the community exists joint action and interaction with other objects. Gravitational fields have the advantage of the joint action of groups on individuals. Modeling the interactions of objects in a gravitational field leads to the self-organized aggregation of objects into clusters. The goal of community discovery is to obtain these clusters. However, the existing researches extracted the description of the basic topology of the network, however, lacked of the field characteristic attributes analysis leads to incomplete modeling information and cannot guarantee the rationality of the community structure division. Community structure has more consistent emotional tendencies considering the emotional factors, which contain stronger influence and cohesion as a crucial character on social media.

The main contributions of this work are as follows:

- The theory of the physical gravitational field is introduced to present the community structure accurately in social networks. The attributes, i.e., the gravitational force between nodes and the internal stability of the gravitational cluster, are integrated into the process of cluster division.

- The set of gravitational kernel nodes is constructed by the random wandering idea, and the adaptive function of the fusion field force extends the cluster to obtain the initial overlapping gravitational field. To further optimize the division of the gravitational field, the overlapping fields can be fused or separated to further obtain a non-overlapping gravitational field.
- Evaluations verify based on four real datasets and five artificial datasets. Comparisons are made with four baseline based on modularity, ARI index, and internal average stability metrics. The experimental results reflect the superiority of the model in this paper.

The remaining of this paper is organized as follows. In Section 2, we discuss the state-of-the-art works related to cluster partitioning and gravitational field modeling. The GF-CDM algorithm is proposed in detail in Section 3. Section 4 compares the GF-CDM algorithm with comparing algorithms for community discovery. The analysis of the evaluation results is also discussed in this section. Section 5 concludes the paper and provides an outlook on possible future directions of work.

## 2 RELATED WORK

In the area of complex network gravitational field research, Li and Du proposed the construction of data fields for two-dimensional static data, and he established the definition of topological potentials and field strengths over the whole space and analyzed the corresponding properties by borrowing the idea of fields in physics [7]. Gan et al. regarded each network node as a field source with a certain quality, and each node acted on other nodes located in the potential field, forming a virtual potential field in the network topological space [8]. Yang et al. proposed a density-based DBSCAN improved clustering method using the average potential difference of each class in the field, which utilized the natural nested structure of equipotential lines (surfaces) in the data potential field to achieve the division of data objects [9]. Ma et al. introduced the gravitational model into the field of influence maximization. The model considered the k-shell value of a node as the node mass and the shortest path between two nodes as the inter-node distance [10]. Li et al. proposed an improved extended gravity model, which was used to identify influential nodes in complex networks [11]. Combining the degree of node size, the model used the information diffusion ability of nodes to characterize the quality of nodes. Li et al. respectively introduced truncated radius and weighted gravity models to reduce high-time complexity [12, 13]. Levy and Goldenberg found that the negative relation between link probability and the inverse of the square of the distance, which is similar to gravity and distance [14]. Wahid-Ul-Ashraf et al. proposed a heuristic link prediction method based on the law of gravity, pointing out that the laws of physics can also be applied to the field of social networks at the local level [15]. Bastami et al. proposed an unsupervised link prediction method based on the gravity model, which significantly reduced the time complexity of the algorithm [16]. Overall, the gravitational field model has many applications in the field of social networks, such as link prediction, community discovery, and influence maximization. Some progress has been made in terms of accuracy and computational complexity.

Using local information to partition communities in complex networks greatly improves efficiency and accuracy. Shang et al. proposed a local community detection method based on high-order structure and edge information (HSEI), focusing on network motif information to select seed nodes, and using the modularity function of the fusion motif to expand the community [4]. Ma et al. adopted a high-order graph clustering method, first identifying the triangular and quadrilateral structures whose connection ratio between internal nodes is tighter than that of external nodes [17]. The structure used local expansion and achieved good performance in high-order graphs. Whang et al. used the k-means algorithm to calculate the multi-level weighted k-kernel of the clusters [18]. Distance function was used to calculate the centroid vertices of each cluster. The neighborhood set of each centroid vertex was used as the seed area of community detection, and then cluster expansion was performed. Veldt et al. proposed an improved flow-based local graph clustering method, which can better combine the semi-supervised information of the target cluster [19]. This method has shown good robustness in experiments. Ding et al. proposed a robust two-stage local community detection algorithm (RTLCD), which selects seed nodes based on node centrality and edge relationship strength [20]. Community members used it as a starting point for community expansion.

The above studies all adopt the cluster partition method based on seed nodes for local expansion. However, some other local division algorithms take a different approach to this. Sheikholeslami and Giannakis developed a top-down method, which regarded the social network graph as composed of some small subgraphs, and used tensors to provide the representation ability of multi-dimensional features of the network [21]. The method improved the quality of the detected communities. The modeling and division of the gravitational field of social networks is proposed based on the cluster division idea of seed expansion. The local fitness method algorithm is the classical algorithm of this type [22]. This method first selected some nodes or some node collections in the network as seed nodes in a certain way. The expansion started from an arbitrary seed node to form a community and stopped when the value of the local fitness function no longer increased until every node was divided. The cluster division method based on seed expansion mainly includes two steps: 1) selection of seed nodes; 2) taking seed nodes as source nodes of community expansion, expanding according to certain function rules, and dividing the network into community sets.

The algorithm based on local expansion assumes that the community is formed around some seed nodes. Therefore selecting seed nodes is a very important step in the local expansion method, which plays a crucial role in the quality of the cluster discovery. Fiala et al. proposed to use the PageRank algorithm to calculate the importance of nodes [23]. The nodes with the largest PageRank value or degree were deleted for community discovery. Whang et al. proposed an algorithm to first sort the nodes in the network by descending order according to the number of neighbors of the nodes, and then extract the top k nodes with the largest number of neighbors as seeds for community discovery [18]. Yang and Zhang calculated the node importance based on the sum of similarities between a node and all its neighbors [24]. The nodes are ranked in descending order and selected sequentially as unvisited nodes for community discovery. Shang et al. find seed nodes by calculating the distribution of similarity communities and obtain neighbouring nodes of similar communities to correct overlapping nodes according to non-central node correction strategy [25]. The corrected overlapping communities are finally obtained. Zhang et al. fused topological similarity and attribute similarity to find seed nodes and performed community expansion based on maximising modularity [26]. For the problem of community discovery in sparse networks, Yue et al. fused the first and second order structure of nodes to select seed nodes to initialise the network, defined the label selection mechanism by combining the neighbouring nodes and the label importance, and updated the labels of nodes [27]. Meng and Liu filtered the community seed nodes by using relative connectivity coefficients between the vertices [28]. The remaining vertices are classified into the same sub-community where the nearest and denser vertices are located.

However, there is a general lack of comprehensive consideration of field properties such as internal stability of the field in the modeling of gravitational fields. The method of selecting seed nodes based on global information has a relatively high time complexity and cannot guarantee the diversity of seeds. To address these issues, we incorporate the attribute of gravitational force between nodes in the seed selection phase and local expansion phase respectively. We propose an algorithm for local expansion community division based on gravity and random wandering ideas.

## 3 METHOD

This chapter first detects overlapping gravitational fields based on inter-nodal gravity and random wandering ideas. Further optimization is done based on the stability of the field itself thus obtaining the non-overlapping gravitational field.

### 3.1 System Framework

To achieve the goal of dividing the emotional gravitational field, we first calculate and compare the magnitude of the forces on nodes by neighbouring fields. This is used to delineate and construct the gravitational field based on the value of the internal stability increment of the gravitational field after the node joins. The algorithm contains two stages: in the first stage, based on the idea of random wandering, the gravitational force between nodes is involved via wandering probability calculation. The N-step transfer probability matrix of nodes is obtained through iteration to get the gravitational seed set. Extend the gravitational field according to the field's force on the nodes to get the division of overlapping gravitational fields. In the second stage, the overlapping fields are fused or separated according to the degree of field overlap and field stability. Further adjustments are made to the attribution of nodes in the overlapping fields to obtain the final delineation of the non-overlapping gravitational fields. The framework of the system is shown in Figure 1.

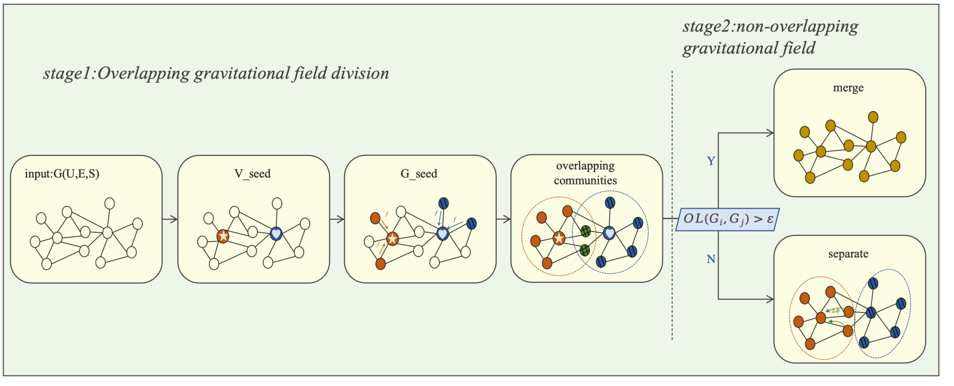



Figure 1. Framework diagram

### 3.2 Node Properties Analysis Based on Gravitational Fields

To model the network structure more comprehensively and accurately, this paper further mines and describes the characteristic properties of the gravitational field and nodes.

#### 3.2.1 Node Property Characteristics

In this paper, three attributes of nodes are considered: node influence, node sentimental tendency value and node quality to fully reflect the importance of nodes in the network. Based on the law of degree centrality [29], taking into account the node's first-order neighbor and second-order neighbor, is expressed in terms of the node's topological importance. The node influence formula is expressed as:

$$\inf(u_i) = \frac{1}{2}(NU_i + \Sigma_{u_j \in NU_i} NU_j), \tag{1}$$

where $NU_i$ denotes the number of nodes $u_i$ the number of neighbouring nodes. Inf denotes the topological importance of a node in the network.

Node sentimental tendency value indicates the sentiment tendency of a user node towards an event, with positive values signify affirmative emotions, whereas negative values imply adverse emotions. Node quality is calculated by combining the topological importance of the user's nodes and the strength of their sentimental tendencies. The larger the node quality, the more important it represents in the gravitational field. It is represented as:

$$m_i = \omega_1 |US_i| + (1 - \omega_1) \inf(v_i), \tag{2}$$

where $w_i$ is the weighting parameter for assigning the node quality impact factor.

### 3.2.2 Characteristics of Gravitational Field Properties

The gravitational field consists of gravitational core nodes and intra-field nodes. The gravitational core node set $V_{seed}$ is the set of nodes with the highest local influence in the field. The intra-field node $G_{k_{in}}$ refers to the node $v_i$ attributed to the gravitational field $G_k$. If node $v_i$ belongs to the gravitational field $G_k$ and has a direct topologically connected edge with the node $u_j$ within the gravitational field $G_j$, then $v_i$ is said to be an inter-field node $G_{k_{between}}$ the fields $G_k$. Inter-field nodes are special intra-field nodes that are bridges connecting several gravitational fields. They can be affected by the forces of several gravitational fields. This subsection focuses on calculating the quality of the gravitational field $M_k$ and its stability $\Phi_k$.

To accurately reflect the characteristics of social networks that include emotional propagation, the attribute of emotional energy $GS_k$ of the gravitational field is proposed. It is the sum of the sentimental tendency values of the intra-field nodes, formula expressed as:

$$GS_k = \sum_{u_i \in G_k} US_i. \tag{3}$$

Assuming the gravitational field $G_1 = \{\nu_1, \nu_2, \nu_3, \nu_4, \nu_5, \nu_6, \nu_7, \nu_8, \nu_9\}$, and letting its sentiment vector be $S_1 = \{0.5, 0.2, 0.3, 0.3, 0.2, -0.3, 0.3, 0.0, -0.5\}$, then the field $G_1$ of sentiment energy denotes 1.0 according to formula (3).

The field quality $M_k$ is used to measure the influence position of the gravitational field in the network. It consists of the value of the gravitational field sentiment energy and the quality of the nodes, and the formula expressed as follows:

$$M_k = |GS_k| + \sum_{u_i \in G_k} m_i. \tag{4}$$

Inspired by Shannon's information entropy formula [30], this paper defines the emotion entropy $\sigma_k$ to measure the degree of sentiment confusion within the gravitational field $G_k$. The smaller its value the higher the sentiment consistency within the field and the more stable the field. The formula is expressed as follows:

$$\begin{aligned} P_{Pk} &= \frac{C_{Pk}}{C_{Pk} + C_{Nk}} \\ P_{Nk} &= \frac{C_{Nk}}{C_{Pk} + C_{Nk}} \\ \sigma_k &= P_{Pk} \log\left(\frac{1}{P_{Pk}}\right) + P_{Nk} \log\left(\frac{1}{P_{Nk}}\right), \end{aligned} \tag{5}$$

where $C_{Pk}$ denotes the sum of positive sentiment tendency values, while $C_{Nk}$ denotes the sum of negative sentiment tendency values. $P_{Pk}$ and $P_{Nk}$ indicate the possibility of positive and negative sentiment occurring during the propagation of emotions, respectively.

The topological tightness of a gravitational field is denoted by $\varphi_k$. The higher value indicates that the nodes in the field are more tightly connected. The equation is expressed as

$$\varphi_k = \frac{\sum_{u_i \in G_k} |e_i|}{|U_k| \cdot (|U_k| - 1)}, \tag{6}$$

where $|e_i|$ denotes a field $G_k$ node $u_i$ the number of contiguous edges with other nodes in the field, and $|U_k|$ denotes the number of edges of the field $G_k$ the number of nodes in the field.

Differences from existing gravitational field models, we denote the field stability by the emotional entropy and the topological stability. It refers to the tightness of the connections within the gravitational field.

$$\Phi_k = \omega_2 \sigma_k + (1 - \omega_2) \varphi_k. \tag{7}$$

### 3.2.3 Characteristics of Force Properties

Not only does the gravitational field exert a force on the nodes external to the field, but the nodes directly connected to it likewise exert a force on it. The inter-node force $f_{ij}$ is given by the formula:

$$f_{ij} = \eta \cdot \frac{m_i \times m_j}{(l_{ij})^2}. \tag{8}$$

$m_i$ and $m_j$ represent the masses of nodes $u_i$ and $u_j$, respectively. $\eta$ represents the polarity of the force between two nodes, which is 1 if they are the same, and $-1$ if they are different. $l_{ij}$ represents the shortest path distance between two nodes.

Here we define $s_{kj}$ to denote the topological distance between the gravitational field $G_k$ and the extra-field node $v_j$:

$$s_{kj} = l_{ij} \left( u_i \in G_k \cap u_i \in NU(\nu_j) \cap \inf(u_i) = \max_{u_m \in NU(\nu_j)} (\inf(u_m)) \right), \tag{9}$$

where $NU(\nu_j)$ denotes the neighborhood of the node $\nu_j$, and $u_i$ is the node in the field $G_k$ with the greatest influence on $\nu_j$.

According to Newton's formula for universal gravitation [31], the force $F_{kj}$ of the gravitational field $G_k$ on the extra-field node $\nu_j$ is computed, which can be expressed as:

$$F_{kj} = \eta \cdot \frac{M_k \times m_j}{(s_{kj})^2}. \tag{10}$$

### 3.3 Gravitational Field Clusters Detection Based on Random Wandering Method of Inter-Node Gravity

Since the community structure is characterized by close connections between nodes in the same community and sparse connections between nodes in different communities. This leads to the basic idea of the random walks idea in community discovery: during random walks, the majority of the wanderer's walks are to nodes in the same community. The probability of wandering between different communities is small. In complex networks, the core node is the most influential node, which is better able to control the dissemination and flow of information. Therefore the choice of core node is important. To obtain the core set of nodes, we first use the inter-node forces as transfer probabilities for directly connected edges in the network. Based on the random walk idea, the $\tau_{step}$ transfer probability matrix $M_N$ is calculated for each node using Markov dynamics. From this a node scoring matrix $SC$ is calculated, the higher the score on this matrix the more connected the node pairs are to each other. Secondly, select the top k nodes in the node scoring matrix as the field core node set $V_{seed}$. For its neighbouring nodes, the inter-node forces are calculated to determine the attribution of the node, thereby constructing the local gravitational cluster $G_{seed}$. For the remaining nodes in the network, we determine their belonging based on the adaptive function fitness. The seed field will be expanded and obtain the overlapping gravitational field division result $G_{initial} = \{G_{in1}, G_{in2}, \ldots, G_{ink}\}$. Finally, the non-overlapping judgment of overlapping emotional gravitational fields is conducted through overlap comparisons.

#### 3.3.1 Selection of Gravitational Field Core Nodes

As shown in Figure 2, pick the node $v_1$ in community $G_1$ as the initial extension node, whose wandering direction has $v_2$, $v_7$, $v_8$ three potential path options. Since the $v_1$, $v_2$ is in the same community, so the node $v_1$ chooses $v_2$ with the highest probability. Compared to the community $G_2$, community $G_1$ is the more attractive to nodes $v_1$. So $v_1$ will choose those nodes in community $G_1$ to wander during this wandering process.

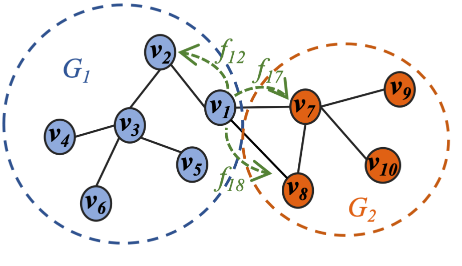



Figure 2. Schematic diagram of a community divided by random wandering

To enhance model performance, normalize the inter-node gravitational force $f_{ij}$ and employ it as a walking probability. Represent the direct transition probabilities between nodes with the transition probability matrix $M$. The initial probability distribution vector of the node $v_i$ is $L_i^1 = (f_{i1}, f_{i2}, \ldots, f_{in})^T$, then the one-step probability transfer matrix is:

$$M = \begin{bmatrix} 0 & v_1 & v_2 & \cdots & v_n \\ v_1 & 0 & f_{12} & \cdots & f_{1n} \\ v_2 & f_{21} & 0 & \cdots & f_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v_n & f_{n1} & f_{n2} & \cdots & 0 \end{bmatrix}, \tag{11}$$

where $f_{ij}$ denotes the force between node $v_i$ and node $v_j$.

The parameter $h$ represents the probability that the walker will remain at the current node. Conversely, $1 - h$ denotes the probability of the walker transitioning to an adjacent node. In this paper, $h$ is set to 0.15. $\tau$ is the threshold of the number of iterations. Based on the six-dimensional space theory [32], this paper sets $\tau$ as 6. The calculation formula of the $\tau_{step}$ probability distribution vector $L_i^\tau$ of node $u_i$ is:

$$L_i^\tau = (1 - h) \times M \times L_i^{\tau-1} + h \times L_i^1. \tag{12}$$

Equation (11) will iterate as the $\tau$ increases until reaches the convergence or the number threshold $\mathcal{E}_1$. The node transfer probability matrix $M_N$ is represented as:

$$M_N = (L_1^T, L_2^T, \ldots, L_n^T)^T. \tag{13}$$

According to the $\tau_{step}$ probability distribution vector $L^\tau$, obtain the rating vector $SC_i$ for node $v_i$:

$$SC_i = \sum_{n=1}^{\tau} L_i^n. \tag{14}$$

Each element $SC_i(\nu_i, \nu_j)$ in the vector $SC_i$ represents the score of the node $v_i$ upon reaching the node $u_j$ after walking $\tau$ steps. Taking into account both global and local information, the model arranges all elements of the node score matrix $SC = (SC_1, SC_2, \ldots, SC_n)$. Select the k core nodes with the largest scores. The selected core nodes become the core of the cluster in the following stage.

### 3.3.2 Expansion of Local Gravitational Clusters

Calculate the attraction of the gravitational core node set to its adjacent nodes according to the field force Equation (8), and the neighbouring nodes affected by the larger gravitational force are added to the field kernel to form a local gravitational cluster $G_{seed}$.

The fitness function can be effective for community discovery. In this paper, the fitness function is improved by the field forces. Divide the nodes into clusters that make the fitness function increase until every node is divided into this cluster. Thus we get the overlapping gravitational field set $G = \{G_1, G_2, \ldots, G_N\}$. The addition of nodes increases the field force in this field, so the improved adaptive fitness function based on the gravitational field is expressed as follows:

$$\text{fitness}\ (G_k) = \frac{F_{in}}{F_{in} + F_{out}}, \tag{15}$$

where $F_{in}$ denotes the cluster $G_k$ the sum of the internal gravitational forces, and $F_{out}$ denotes the sum of the $G_k$ the sum of the external gravitational forces.

### 3.4 Overlapping Gravitational Field Optimization Algorithm for Fusion Field Stability

As the cluster expands, a node may belong to multiple gravitational fields. Since overlapping field is not conducive to the study of group affective convergence, the following two aspects need to be carried out: (i) to judge the fusion or separation of overlapping fields; (ii) to distinguish the field affiliation of nodes in the overlapping fields that need to be separated.

We use Equation (16) in this paper to calculate the field overlapping degree, where $|E\left(G_i\right)|$ denotes the number of nodes in the field $G_i$.

$$OL\left(G_i, G_j\right) = \frac{|E\left(G_i\right) \cap E\left(G_j\right)|}{|E\left(G_i\right)| + |E\left(G_j\right)|}. \tag{16}$$

We define a set of overlapping nodes, $OLG$, and add nodes belonging to more than one field to the set, $OLG = \{ol_1, ol_2, \ldots, ol_n\}$.

The field delineation and the attribution of overlapping nodes reflect the degree of overlap and field stability:

- If $OL\left(G_i, G_j\right) \geq \varepsilon$, where shows that the fields $G_i$ and $G_j$ have a high enough overlap. The algorithm merges two fields $(G_i, G_j)$ and forms a new gravitation field.
- If $OL(G_i, G_j) < \varepsilon$, the fields $G_i$ and $G_j$ need to be separated, as well as the overlapping nodes, which need to be reasonably divided by one of the field uniquely. The decision on overlapping nodes' attribution depends on the following rules: this algorithm uses the increment of the field stability itself as the main factor in determining the belonging of the nodes. Based on the method, the node is considered to be beneficial to the optimization of the community structure, when the addition of a node outside the field will make the increment of field stability greater than a threshold value. Therefore, this algorithm will calculate separately the effect of overlapping nodes $ol_i$ on the stability of the field $G_{ij}$ stability and select the node with the largest incremental field stability $\Delta\Phi$. The field with the largest incremental field stability is selected as the unique attribution of the node.

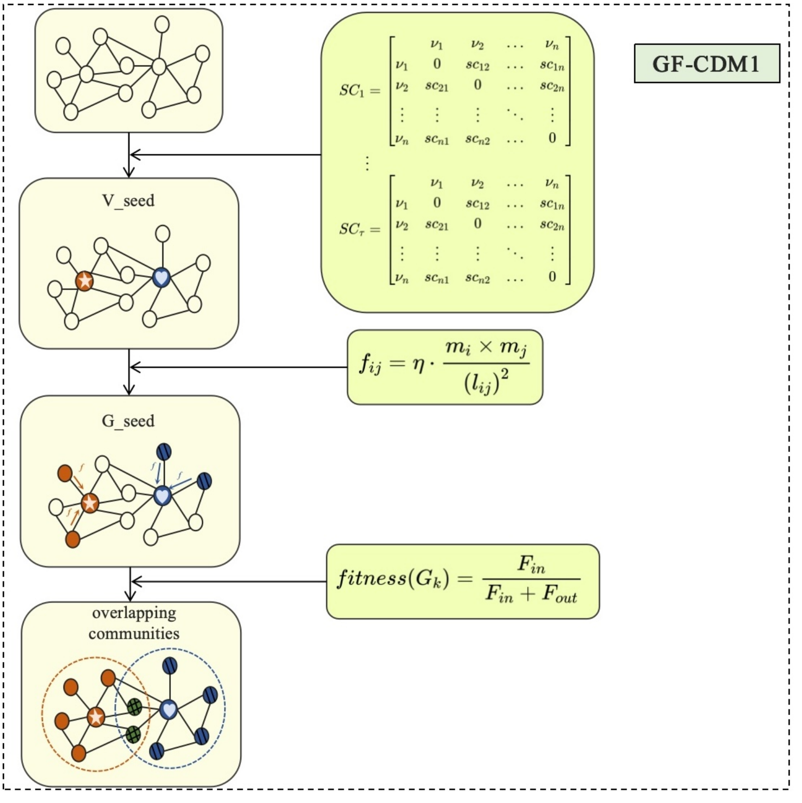



Figure 3. Flowchart of GF-CDM1

**GF-CDM1:** First, the social network model $(U, S, E)$ is established. The node transfer probability matrix MM is constructed using the inter-node force of gravity based on the random wandering thought. Thus, the node scoring matrix $SC = (SC_1, SC_2, \dots, SC_n)$ is calculated to select the top k nodes as seed nodes. The local gravitational cluster is extended using the adaptive function fitness of the fusion field force values, and the overlapping gravitational field division results are output $G_{in} = \{G_{in1}, G_{in2}, \dots, G_{inN}\}$. The algorithm for this phase is divided into two sub-algorithms, i.e. The core node selection algorithm and the cluster expansion algorithm, whose pseudo-codes are shown as Algorithm 1 and Algorithm 2.

**GF-CDM2:** By comparing the degree of field overlap with a predefined threshold, separation or fusion operations are applied to overlapping fields based on the results. The output is a non-overlapping gravitational field cluster. The specific algorithm for this phase is in Algorithm 3.

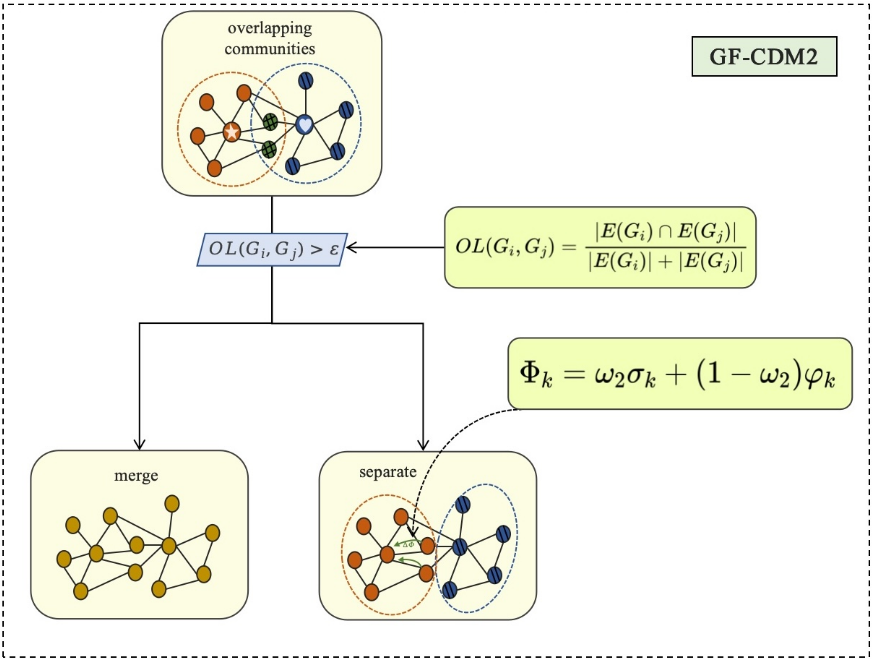



Figure 4. Flowchart of GF-CDM2

### 3.5 Complexity Analysis

Assuming the graph has $n$ nodes. In Algorithm 1, lines 3–8 calculate the gravitational magnitude and one-step transfer probability matrix between nodes in the network with a time complexity of $O(a \cdot n)$. Here $a$ is the neighborhood size of a single node, $a \ll n$. Lines 9–17 iteratively compute $\tau$ step transfer probability matrix $M_N$ and obtain the node scoring matrix $SC_i$ with time complexity is $O(\tau \cdot n^2)$. Here $\tau$ is the number of iterations with $\tau \ll n$. Based on the node scoring matrix $SC$, the set of field core nodes is selected $V_{seed}$ with time complexity of $O(a \cdot k)$ as lines 18–27, where $k$ is the threshold of the number of field core nodes. Therefore, the total time complexity of the field core node selection algorithm via inter-node gravity and random wandering is $O(\tau \cdot n^2 + a \cdot (n + k))$, which is approximated to $O(\tau \cdot n^2)$.

In Algorithm 2, lines 3–12 construct the local gravitational group set $G_{seed}$ with a time complexity of $O(a \cdot k)$, where $a$ is the neighborhood size of the field core nodes, and $k$ is the number of field core nodes. Lines 13–22 extend the local gravitational field using the adaptive function fitness with a time complexity of $O(k \cdot b)$, where $b$ is the neighborhood size of the local gravitational cluster. Lines 23–29 determine the field attribution of free nodes, with time complexity $O\left(k \cdot (n - c)\right)$. $n$ is the number of network nodes, and $c$ is the size of the local gravitational cluster with

**Algorithm 1:** Gravitational field core node selection algorithm

**Input:** Network diagram $H = (U, E, S)$, Neighborhood $N_U$, parameters $\omega_1$, parameters $k$

**Output:** Gravitational field core node set $V_{seed}$

Initialize the node transfer probability matrix $M_N$, node scoring matrix $SC_i$, Field kernel node set $V_{seed}$ ;
**for** $i = 1$ *to* $|E|$ **do**
    **for** $j = 1$ *to* $|NU_i|$ **do**
        Calculate the nodes according to Equation (9) $u_i$ and $u_j$ the gravitational force between $f_{ij}$. $u_j \in NU_i$ ;
    Normalize $f_{ij}$ After normalization, a one-step transfer probability vector is obtained $L_i^1$ ;
**for** $i = 1$ *to* $|E|$ **do**
    **for** $\tau = 2$ *to* $\epsilon_1$ **do**
        $L_i^\tau = (1 - h) \times M \times L_i^{\tau-1} + h \times m_i$ ;
        $SC_i = SC_i + L_i^\tau$ ;
        **if** $L_i^\tau == L_i^{\tau-1}$ **then**
            **break** ;
Score Node Scoring Matrix $SC = (SC_1, SC_2, \ldots, SC_n)$ ;
**for** *nodes in* $U$ **do**
    **if** $SC_i > SC_{NU_i}$ **then**
        $v_i \in V_{seed}$ ;
        $U = U - v_i - \tau - NU_i$ ;
        $cnt = cnt + 1$ ;
    **if** $U = \emptyset$ *or* $cnt = k$ **then**
        **break** ;
**return** $V_{seed}$ ;

$a, k \ll n$. Therefore, the total time complexity of the overlapping gravitational field cluster detection algorithm is $O(k \cdot (n + a + b - c))$, which is approximated by $O(k{\cdot}n)$.

In Algorithm 3, lines 3–15 merge two fields with overlapping degrees over the threshold value with a time complexity of $O(k{\cdot}k)$, where $k$ denotes the number of fields. Lines 16–26 determine the attribution of the field to the remaining nodes in the set of overlapping nodes. Choose the field with the largest incremental stability $\Delta\Phi$ with a time complexity of $O(a{\cdot}n)$, where $n$ denotes the number of remaining nodes, $a$ denotes the number of nodal neighbouring fields. $a, k \ll n$. Therefore, the total time complexity of the non-overlapping gravitational field cluster detection method is $O(a{\cdot}\, n + k{\cdot}\, k)$, which is approximated by $O(a{\cdot}n)$.

**Algorithm 2:** Overlapping Gravitational Field Partitioning Algorithm

**Input:** Network diagram $H = (U, E, S^{t_0})$, Field Node Set $V_{seed}$
**Output:** Overlapping gravitational field division $G_{initial} = \{G_{in1}, G_{in2}, \ldots, G_{ink}\}$

Initialize local gravitational cluster sets $G_{seed} = \{V_{seed}\}$, the set of free nodes $UF = U - V_{seed}$;
**for** $i = 1$ *to* $k$ **do**
  **for** $j = 1$ *to* $NU_i$ **do**
    Calculate the gravity $f_{ij}$ between the field core node $seed_i$ and the node $u_j$ according to Equation (8), $u_j \in NU_i$;
    **if** $(f_{ij} \geq \varepsilon_2)$ **then**
      Adding nodes $u_j$ to local gravity group $G_{seed_i}$;
Local gravitational clusters $G_{seed} = \{G_{seed_1}, G_{seed_2}, \ldots, G_{seed_k}\}$, and Free Node Set $UF = U - G_{seed}$;
**for** $i = 1$ *to* $k$ **do**
  **for** $u_j \in NU_{G_{seed_i}}$ **do**
    **if** $(fitness(G_{seed_i} + u_j) > fitness(G_{seed_i}))$ **then**
      Adding nodes $u_j$ to local gravity group $G_{seed_i}$;
      **if** $(u_j \in UF)$ **then**
        $UF = U - u_j$
**while** $(UF \neq \emptyset)$ **do**
  **for** $u_j \in UF, \quad G_{seed_k} \in NU_j$ **do**
    $\Delta fitness(u_j) = fitness(G_{seed_k} + u_j) - fitness(G_{seed_k})$;
    Select the field with the largest $\Delta fitness$ to join;
    $UF = UF - u_j$
**return** $G_{initial} = \{G_{in1}, G_{in2}, \ldots, G_{ink}\}$

## 4 EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we evaluate the performance with experiments on four real and five synthetic network datasets to verify the effectiveness of the proposed GF-CDM. The processor used in the experiment is a dual-core Intel Core i5 running at 1.8 GHz, and the memory is 8.0 GB 1 600 MHz DDR3.

### 4.1 Experimental Dataset

To evaluate the performance of the proposed GF-CDM1 and GF-CDM2 cluster detection models, we experiment on real and synthetic network datasets. The average value obtained from 10 runs of each simulation is used as the final result.

**Algorithm 3:** Non-overlapping gravitational field cluster detection algorithm

**Input:** Overlapping gravitational field division $G_{\text{initial}} = \{G_{\text{in}1}, G_{\text{in}2}, \ldots, G_{\text{in}k}\}$, set of overlapping fields OLG, overlap threshold parameter $\varepsilon$

**Output:** Non-overlapping gravitational field division $G = \{G_1, G_2, \ldots, G_N\}$

Initialize the set of overlapping nodes $\text{OLG} = \{\text{ol}_1, \text{ol}_2, \ldots, \text{ol}_n\}$, the set of overlapping node neighborhood fields $\text{NU}_G$ ;
**for** $i = 1$ *to* $k$ **do**
  **for** $j = i$ *to* $k$ **do**
    **if** $G_{ini} \cap G_{inj} \neq \emptyset$ **then**
      Calculate the field overlap according to Equation (16) $\text{OL}(G_{\text{in}i}, G_{\text{in}j})$ ;
      Calculate the field stability according to Equation (7) $\Phi_{\text{in}i,\text{in}j}$ ;
      **if** $OL(G_{ini}, G_{inj}) \geq \varepsilon$ **then**
        $G_i = G_{\text{in}i} + G_{\text{in}j} - \text{OLG}_{\text{in}i,\text{in}j}$ ;
        $G_{\text{in}i} = G_i$ ;
        $\text{OLG} = \text{OLG} - \text{OLG}_{\text{in}i,\text{in}j}$ ;
**while** $OLG \neq \emptyset$ **do**
  **for** $i = 1$ *to* $|OLG|$ **do**
    **for** $j = 1$ *to* $|NU_G|$ **do**
      Calculated from Equation (7) $\Delta\Phi_j = \Phi_{\text{Gnu}-\text{ol}_i} - \Phi_{\text{Gnu}}$ ;
      $k = \arg\max \Delta\Phi_j$ ;
    $G_k = G_k + \text{ol}_i$ ;
    $\text{OLG} = \text{OLG} - \text{ol}_i$ ;
**return** $G = \{G_1, G_2, \ldots, G_N\}$

1. Real network dataset

- **Dolphins** [33], the Dolphin Network, is a non-directional social network formed by regular communication between 62 dolphins living in New Zealand;
- **Email-univ** [34], which is the topology of an email link between university colleges;
- **Amherst41** [35], which is a network of social friends extracted from Facebook;
- **Advogato** [36], which is a social community platform, where user can explicitly express their weighted trust relationships with each other.

The details of 4 real databases are listed in Table 1.

| Dataset | Number of Nodes | Number of Consecutive Edges | Average Clustering Coefficient |
|---|---|---|---|
| Dolphins | 62 | 159 | 0.2590 |
| Email-univ | 1 133 | 5 451 | 0.2202 |
| Amherst41 | 2 235 | 90 954 | 0.3104 |
| Advogato | 6 551 | 51 332 | 0.2868 |

Table 1. Basic information on the real data set

2. Synthetic network dataset

   Due to the good representation of node degree and community size heterogeneity, LFR-benchmark is widely used for generating synthetic networks [37]. In this experiment, five sets of artificially simulated networks were generated using the LFR-benchmark program, their vertex numbers are 1 000, 2 000, 3 000, 4 000 and 5 000. The basic information of each LFR network is shown in Table 2. The parameter $\mu$ denotes the mixing parameter. The higher the $\mu$ of the network, the more dispersed the community is; the smaller the $\mu$, the more compact the community is. In the experiments of this paper, it is set to 0.3.

| Parameter Name | Description | Value |
|---|---|---|
| $N$ | Number of nodes | $[1\,000, 5\,000]$ |
| $\mu$ | Mixing parameters | 0.3 |
| k | Average node degree | 20 |
| $k_{\max}$ | Maximum node degree | 50 |
| $C_{\max}$ | Maximum number of community nodes | 100 |
| $C_{\min}$ | Minimum number of community nodes | 20 |

Table 2. Basic information on the LFR artificial network dataset

## 4.2 Comparison Algorithm

The experiments compared with the following 4 classical community detection algorithms.

- **FastQ** [38] is a fast modularity optimization algorithm. It merges clusters continuously depending on the largest increment and the smallest decrement of modularity to find the optimal graph partition with the maximum modularity.
- **LFM** [22] is a locally extended overlapping cluster detection algorithm based on the definition of an adaptive degree function fitness, where a community consists of a set of nodes that maximize fitness.
- **LPA** [39] is a label propagation algorithm, the nodes in the network are initialized with different labels, randomly ordered and updated label as the majority label of its nearest neighbors.

- **WalkTrap** [40] is a random walk algorithm that uses a finite steps random walk process on a network to calculate the probability of one point reaching another point in the network. The similarity between the two points is also analyzed, using hierarchical clusters with different levels.

### 4.3 Evaluation Indicators

This experiment uses modularity, ARI index, and internal average stability as metrics for the evaluations.

1. Modularity

   Modularity is to measure the effectiveness of community segmentation. When the similarity of nodes within a community is relatively high and the similarity of nodes outside the community is relatively low, it is considered a more desirable community detection result. The formula for calculating the degree of modularity is as follows:

$$Q = \sum_{C} \left[ \frac{\sum in}{2m} - \left( \frac{\sum out}{2m} \right)^2 \right] = \sum_{C} \left[ e_c - a_c^2 \right], \tag{17}$$

   where $\sum in$ denotes the sum of inter-node forces within the community $c_i$, and $\sum out$ denotes the sum of the external force connected to the community $c_i$ denotes the sum of the external edge weights connected to the community, and $m$ denotes the sum of the number of edges in all communities, and the higher the value of the modularity, the better the community detection.

2. ARI Index

   ARI is used to evaluate the effectiveness of clustering that denotes:

$$ARI = \frac{\frac{(a_{11}+a_{01})(a_{11}+a_{10})}{a_{00}}}{\frac{(a_{11}+a_{01})+(a_{11}+a_{10})}{2} - \frac{(a_{11}+a_{01})(a_{11}+a_{10})}{a_{00}}}, \tag{18}$$

   where $a_{11}$ denotes the number of pairs of nodes belonging to the same community in both the real and experimental community divisions. $a_{00}$ is the number of pairs of nodes that do not belong to the same community in both the real and the experimental community divisions. $a_{10}$ is the number of pairs of nodes that belong to the same community in the real community but not in the experimental community. $a_{01}$ denotes the number of pairs of nodes that do not belong to the same community in the real community but belong to the same community in the experiment.

3. Internal average stability

   $\varphi$ is used to measure the average stability within a community. Higher values indicate stronger connections between nodes within the community. The formula is expressed as:

$$\varphi = \sum_{c} \frac{\sum_{u_i \in C_k} |e_i|}{|U_k| \cdot (|U_k| - 1)} / K, \tag{19}$$

where $|e_i|$ denotes a field $G_k$ node $u_i$ the number of contiguous edges with other nodes in the field, and $|U_k|$ denotes the number of edges of the field $G_k$ the number of nodes within the field, and $K$ denotes the total number of network communities obtained by the algorithm division.

### 4.4 Experimental Results

In this section, we will analyze the experimental results. We have chosen modularity and average stability as the evaluation metrics for real network partitions. As real network datasets often lack clear ground truth community annotations, we employ the Adjusted Rand Index (ARI) as the evaluation metric for synthetic network partitions.

#### 4.4.1 Real Dataset Experimental Results

1. Performances of modularity

   As shown in Table 3, the experiment was conducted on four real network datasets and the results of the communities classified by each algorithm were evaluated using the modularity metric.

| | FastQ | LFM | LPA | WalkTrap | GF-CDM1 | GF-CDM2 |
|---|---|---|---|---|---|---|
| Dolphins | 0.5048 | 0.3656 | 0.3804 | 0.5137 | 0.4955 | 0.5004 |
| Email-univ | 0.4847 | 0.4365 | 0.4342 | 0.4866 | 0.4815 | 0.4923 |
| Advogato | 0.3368 | 0.2883 | 0.0537 | 0.3123 | 0.3434 | 0.3375 |
| Amherst41 | 0.3669 | 0.3304 | 0.3120 | 0.3745 | 0.3677 | 0.3750 |

Table 3. Performances of modularity metrics

From Figure 5, it can be observed that our method performs relatively well, especially when dealing with a larger number of communities. In most cases, the results of GF-CDM2 are superior to GF-CDM1. This improvement in modularity is due to GF-CDM2's optimization for non-overlapping community partitioning based on field stability, building upon GF-CDM1. When the number of communities is small, WalkTrap is an algorithm that terminates with global maximization of modularity. Although our method is intermediate, the difference compared to the optimal results is minimal. Our approach is based on the LFM algorithm, and the results indicate that our method outperforms LFM in terms of performance. This demonstrates that the proposed algorithm is based on iterative random walk algorithms to obtain a probability matrix, from which a node score matrix is derived. The improvement in field kernel node selection based on this method is effective.

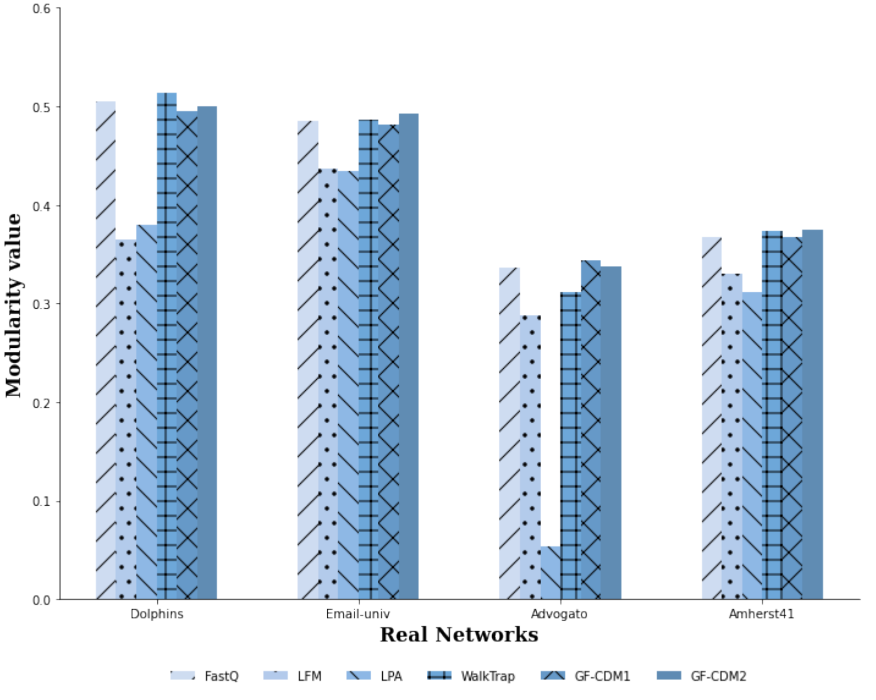



Figure 5. Comparison of modularity metrics within each algorithmic community

2. Comparison of average internal stability

This experiment was conducted on four real network datasets, using internal average stability as an evaluation metric.

The analysis of Figure 6 indicates that the GF-CDM algorithm proposed in this study, when applied to the Dolphins dataset, exhibits a slightly lower average stability within communities compared to the WalkTrap algorithm. However, on other real networks, the modularity of GF-CDM surpasses that of other algorithms. For datasets with a smaller number of nodes, the WalkTrap algorithm, based on random walks, performs well, with minimal differences compared to the algorithm proposed in this study. Among the six algorithms, their performance generally declines on the Advogato dataset. This is primarily attributed to the less pronounced community structure of the dataset, making it challenging to achieve effective community partitioning. The proposed algorithm, in its second stage, optimizes community partitioning through the fusion of field forces and field stability, resulting in a significant improvement in the average stability within communities.

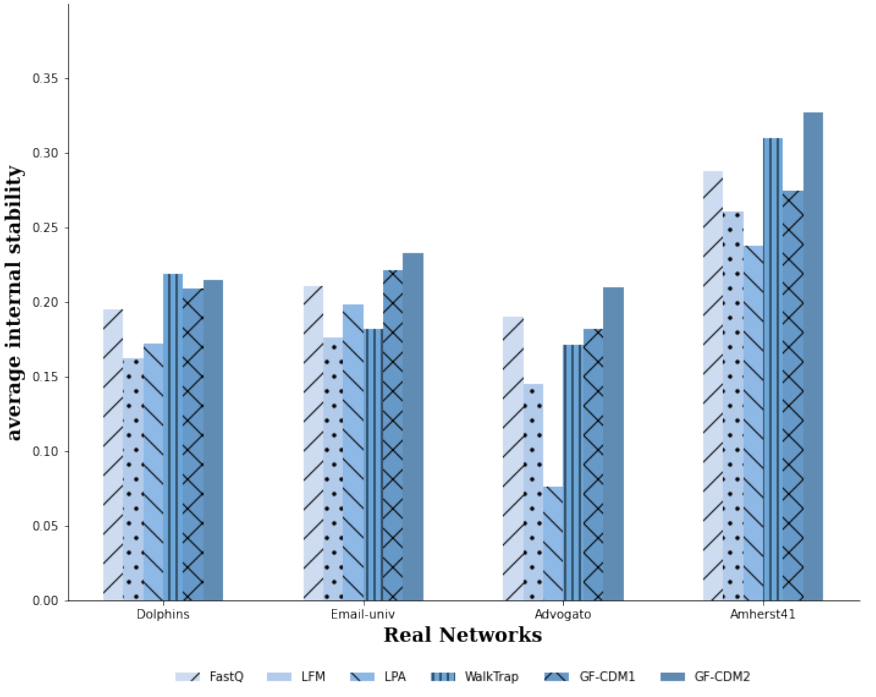



Figure 6. Comparison of average stability metrics within each algorithmic community

### 4.4.2 Synthetic Network Experimental Results

This experiment was conducted in five artificial networks generated by the LFR-Benchmark program, and the community segmentation results of each algorithm were compared with the community labels in the artificial network dataset for the ARI metrics, and the experimental results are shown in Figure 7.

Analysis of Figure 7 reveals that, with the increase in network size, the Adjusted Rand Index (ARI) of all six algorithms shows a declining trend. However, the proposed GF-CDM algorithm generally maintains a relatively stable level. This is attributed to the GF-CDM algorithm's utilization of sentiment field forces for local expansion, coupled with the consideration of the superiority of group interactions over individual interactions, thus enhancing the quality of community discovery. At a node count of 3 000, the performance of the proposed algorithm exhibits a slight decrease compared to FastQ and WalkTrap algorithms, although the difference is marginal. In contrast to the LFM algorithm, the proposed algorithm demonstrates a significant improvement, validating the effectiveness of the improvement strategy that comprehensively considers both global and local information during the selection of core nodes in the field and the cluster expansion phase. Consequently, this enhances the quality of clusters obtained by the partitioning algorithm.

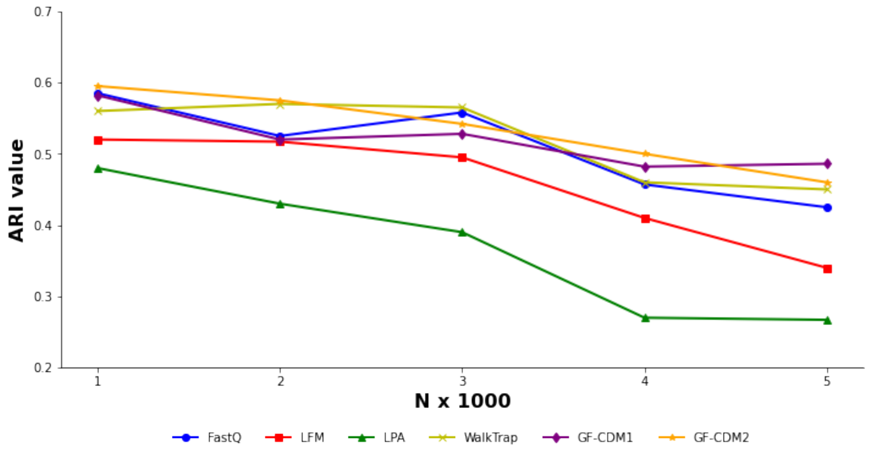



Figure 7. Experimental results of each algorithm in synthetic networks of different sizes

## 4.5 Discussion

To validate the superior performance of our model in community partitioning, we conducted experiments covering four real network datasets and five artificially generated datasets with varying scales. We compared the experimental results with those of four benchmark algorithms. The results significantly demonstrate the excellence of our model in terms of modularity, ARI index, and average stability compared to the benchmark algorithms. This superiority stems from our comprehensive consideration of both global and local information during the selection of field core nodes and the expansion phase of community clusters in the algorithm. In the second stage, we optimized community partitioning by integrating field forces and field stability. The successful implementation of these improvement strategies underscores our meticulous algorithm design, ensuring the consistency of results.

# 5 CONCLUSION

In this paper, we introduce the gravitational field theory in physics and propose a new gravitational field modeling and partitioning method named GF-CDM. First, it defines the various characteristics of nodes and gravitational fields that contain emotional properties. Two stages of field model partitioning realize the overlapping and non-overlapping clusters. The first stage depends on inter-node gravitational force and random wandering idea, which extends local overlapping cluster discovery. Firstly, we determine the transfer probability matrix through the random walk idea. The node scoring matrix is calculated according to the transfer probability matrix to construct the gravitational core node set. The local gravitational cluster is extended according to the adaptive function of the fusion field force to obtain the overlapping gravitational field cluster. The second stage fuses field stability to optimize overlapping gravitational fields. Based on the degree of overlap, the resulting overlapping fields are fused or separated. The belonging calculation is performed on the overlapping nodes. Non-overlapping gravitational field is obtained. The proposed model GF-CDM is compared to four benchmark algorithms in terms of modularity, ARI index, and internal average stability. This model outperforms the other baselines and is more interpretable. In future work, adequate modeling of user node characteristics of real networks can be considered to achieve better partitioning results.

## REFERENCES

[1] Djerbi, R.—Imache, R.—Amad, M.: Communities' Detection in Social Networks: State of the Art and Perspectives. 2018 International Symposium on Networks, Computers and Communications (ISNCC), IEEE, 2018, pp. 1–6, doi: 10.1109/ISNCC.2018.8531055.

[2] Girvan, M.—Newman, M. E. J.: Community Structure in Social and Biological Networks. Proceedings of the National Academy of Sciences of the United States of America (PNAS), Vol. 99, 2002, No. 12, pp. 7821–7826, doi: 10.1073/pnas.122653799.

[3] Zhang, W.—Shang, R.—Jiao, L.: Complex Network Graph Embedding Method Based on Shortest Path and MOEA/D for Community Detection. Applied Soft Computing, Vol. 97, Part A, 2020, Art. No. 106764, doi: 10.1016/j.asoc.2020.106764.

[4] Shang, R.—Zhang, W.—Zhang, J.—Feng, J.—Jiao, L.: Local Community Detection Based on Higher-Order Structure and Edge Information. Physica A: Statistical Mechanics and Its Applications, Vol. 587, 2022, Art. No. 126513, doi: 10.1016/j.physa.2021.126513.

[5] Havemann, F.—Heinz, M.—Struck, A.—Gläser, J.: Identification of Overlapping Communities and Their Hierarchy by Locally Calculating Community-Changing Resolution Levels. Journal of Statistical Mechanics: Theory and Experiment, Vol. 2011, 2011, No. 1, Art. No. P01023, doi: 10.1088/1742-5468/2011/01/P01023.

[6] Hu, F.—Liu, Y.: A New Algorithm CNM-Centrality of Detecting Communities Based on Node Centrality. Physica A: Statistical Mechanics and Its Applications, Vol. 446, 2016, pp. 138–151, doi: 10.1016/j.physa.2015.10.083.

[7] Li, D.—Du, Y.: Artificial Intelligence with Uncertainty. CRC Press, 2017, doi: 10.1201/9781315366951.

[8] Gan, W. Y.—He, N.—Li, D. Y.—Wang, J. M.: Community Discovery Method in Networks Based on Topological Potential. Journal of Software, Vol. 20, 2009, No. 8, pp. 2241–2254 (in Chinese).

[9] Yang, J.—Gao, J.—Liang, J.—Liu, Y.: An Improved DBSCAN Clustering Algorithm Based on Data Field. Journal of Frontiers of Computer Science & Technology, Vol. 6, 2012, No. 10, pp. 903–911 (in Chinese).

[10] Ma, L. L.—Ma, C.—Zhang, H. F.—Wang, B. H.: Identifying Influential Spreaders in Complex Networks Based on Gravity Formula. Physica A: Statistical Mechanics and Its Applications, Vol. 451, 2016, pp. 205–212, doi: 10.1016/j.physa.2015.12.162.

[11] Li, H.—Shang, Q.—Deng, Y.: A Generalized Gravity Model for Influential Spreaders Identification in Complex Networks. Chaos, Solitons & Fractals, Vol. 143, 2021, Art. No. 110456, doi: 10.1016/j.chaos.2020.110456.

[12] Li, Z.—Ren, T.—Ma, X.—Liu, S.—Zhang, Y.—Zhou, T.: Identifying Influential Spreaders by Gravity Model. Scientific Reports, Vol. 9, 2019, No. 1, Art. No. 8387, doi: 10.1038/s41598-019-44930-9.

[13] Liu, F.—Wang, Z.—Deng, Y.: GMM: A Generalized Mechanics Model for Identifying the Importance of Nodes in Complex Networks. Knowledge-Based Systems, Vol. 193, 2020, Art. No. 105464, doi: 10.1016/j.knosys.2019.105464.

[14] Levy, M.—Goldenberg, J.: The Gravitational Law of Social Interaction. Physica A: Statistical Mechanics and Its Applications, Vol. 393, 2014, pp. 418–426, doi: 10.1016/j.physa.2013.08.067.

[15] Wahid-Ul-Ashraf, A.—Budka, M.—Musial, K.: How to Predict Social Relationships – Physics-Inspired Approach to Link Prediction. Physica A: Statistical Mechanics and Its Applications, Vol. 523, 2019, pp. 1110–1129, doi: 10.1016/j.physa.2019.04.246.

[16] Bastami, E.—Mahabadi, A.—Taghizadeh, E.: A Gravitation-Based Link Prediction Approach in Social Networks. Swarm and Evolutionary Computation, Vol. 44, 2019, pp. 176–186, doi: 10.1016/j.swevo.2018.03.001.

[17] Ma, W.—Cai, L.—He, T.—Chen, L.—Cao, Z.—Li, R.: Local Expansion and Optimization for Higher-Order Graph Clustering. IEEE Internet of Things Journal, Vol. 6, 2019, No. 5, pp. 8702–8713, doi: 10.1109/jiot.2019.2923228.

[18] Whang, J. J.—Gleich, D. F.—Dhillon, I. S.: Overlapping Community Detection Using Neighborhood-Inflated Seed Expansion. IEEE Transactions on Knowledge and Data Engineering, Vol. 28, 2016, No. 5, pp. 1272–1284, doi: 10.1109/tkde.2016.2518687.

[19] Veldt, N.—Klymko, C.—Gleich, D. F.: Flow-Based Local Graph Clustering with Better Seed Set Inclusion. Proceedings of the 2019 SIAM International Conference on Data Mining, 2019, pp. 378–386, doi: 10.1137/1.9781611975673.43.

[20] Ding, X.—Zhang, J.—Yang, J.: A Robust Two-Stage Algorithm for Local Community Detection. Knowledge-Based Systems, Vol. 152, 2018, pp. 188–199, doi: 10.1016/j.knosys.2018.04.018.

[21] Sheikholeslami, F.—Giannakis, G. B.: Overlapping Community Detection via Constrained PARAFAC: A Divide and Conquer Approach. 2017 IEEE International Conference on Data Mining (ICDM), 2017, pp. 127–136, doi: 10.1109/ICDM.2017.22.

[22] Lancichinetti, A.—Fortunato, S.—Kertész, J.: Detecting the Overlapping and Hierarchical Community Structure in Complex Networks. New Journal of Physics, Vol. 11, 2009, No. 3, Art. No. 033015, doi: 10.1088/1367-2630/11/3/033015.

[23] Fiala, D.—Rousselot, F.—Ježek, K.: PageRank for Bibliographic Networks. Scientometrics, Vol. 76, 2008, No. 1, pp. 135–158, doi: 10.1007/s11192-007-1908-4.

[24] Yang, J. X.—Zhang, X. D.: Finding Overlapping Communities Using Seed Set. Physica A: Statistical Mechanics and Its Applications, Vol. 467, 2017, pp. 96–106, doi: 10.1016/j.physa.2016.10.006.

[25] Shang, R.—Zhao, K.—Zhang, W.—Feng, J.—Li, Y.—Jiao, L.: Evolutionary Multiobjective Overlapping Community Detection Based on Similarity Matrix and Node Correction. Applied Soft Computing, Vol. 127, 2022, Art. No. 109397, doi: 10.1016/j.asoc.2022.109397.

[26] Zhang, W.—Zhao, K.—Shang, R.: Evolutionary Multi-Objective Attribute Community Detection Based on Similarity Fusion Strategy with Central Nodes. Applied Soft Computing, Vol. 150, 2024, Art. No. 111101, doi: 10.1016/j.asoc.2023.111101.

[27] Yue, Y.—Wang, G.—Hu, J.—Li, Y.: An Improved Label Propagation Algorithm Based on Community Core Node and Label Importance for Community Detection in Sparse Network. Applied Intelligence, Vol. 53, 2023, No. 14, pp. 17935–17951, doi: 10.1007/s10489-022-04397-0.

[28] Meng, Y.—Liu, X.: Finding Central Vertices and Community Structure via Extended Density Peaks-Based Clustering. Information, Vol. 12, 2021, No. 12, Art. No. 501, doi: 10.3390/info12120501.

[29] Chen, Q.—Wu, T. T.: A Method for Local Community Detection by Finding Maximal-Degree Nodes. 2010 International Conference on Machine Learning and Cybernetics, Vol. 1, 2010, pp. 8–13, doi: 10.1109/ICMLC.2010.5581103.

[30] Shannon, C. E.: A Mathematical Theory of Communication. The Bell System Technical Journal, Vol. 27, 1948, No. 3, pp. 379–423, doi: 10.1002/j.1538-7305.1948.tb01338.x.

[31] Newton, I.—Cohen, I. B.—Whitman, A.: The Principia: Mathematical Principles of Natural Philosophy. University of California Press, 1999.

[32] Weinberg, S.: Six-Dimensional Methods for Four-Dimensional Conformal Field Theories. Physical Review D – Particles, Fields, Gravitation, and Cosmology, Vol. 82, 2010, No. 4, Art. No. 045031, doi: 10.1103/PhysRevD.82.045031.

[33] Lusseau, D.—Schneider, K.—Boisseau, O. J.—Haase, P.—Slooten, E.—Dawson, S. M.: The Bottlenose Dolphin Community of Doubtful Sound Features a Large Proportion of Long-Lasting Associations: Can Geographic Isolation Explain This Unique Trait? Behavioral Ecology and Sociobiology, Vol. 54, 2003, No. 4, pp. 396–405, doi: 10.1007/s00265-003-0651-y.

[34] Traud, A. L.—Mucha, P. J.—Porter, M. A.: Social Structure of Facebook Networks. Physica A: Statistical Mechanics and Its Applications, Vol. 391, 2012, No. 16, pp. 4165–4180, doi: 10.1016/j.physa.2011.12.021.

[35] Red, V.—Kelsic, E. D.—Mucha, P. J.—Porter, M. A.: Comparing Community Structure to Characteristics in Online Collegiate Social Networks. SIAM Review, Vol. 53, 2011, No. 3, pp. 526–543, doi: 10.1137/080734315.

[36] Massa, P.—Salvetti, M.—Tomasoni, D.: Bowling Alone and Trust Decline in Social Network Sites. 2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing, 2009, pp. 658–663, doi: 10.1109/DASC.2009.130.

[37] Lancichinetti, A.—Fortunato, S.—Radicchi, F.: Benchmark Graphs for Testing Community Detection Algorithms. Physical Review E – Statistical, Nonlinear,

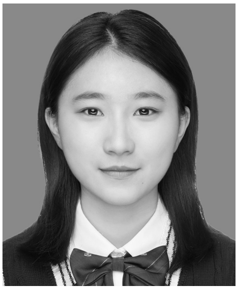
and Soft Matter Physics, Vol. 78, 2008, No. 4, Art. No. 046110, doi: 10.1103/PhysRevE.78.046110.

[38] Clauset, A.—Newman, M. E. J.—Moore, C.: Finding Community Structure in Very Large Networks. Physical Review E – Statistical, Nonlinear, and Soft Matter Physics, Vol. 70, 2004, No. 6, Art. No. 066111, doi: 10.1103/PhysRevE.70.066111.

[39] Raghavan, U. N.—Albert, R.—Kumara, S.: Near Linear Time Algorithm to Detect Community Structures in Large-Scale Networks. Physical Review E – Statistical, Nonlinear, and Soft Matter Physics, Vol. 76, 2007, No. 3, Art. No. 036106, doi: 10.1103/PhysRevE.76.036106.

[40] Pons, P.—Latapy, M.: Computing Communities in Large Networks Using Random Walks. In: Yolum, P., Güngör, T., Gürgen, F., Özturan, C. (Eds.): Computer and Information Sciences – ISCIS 2005. Springer, Berlin, Heidelberg, Lecture Notes in Computer Science, Vol. 3733, 2005, pp. 284–293, doi: 10.1007/11569596_31.

**Meizi Li** is currently Associate Professor at the College of Information, Mechanical and Electrical Engineering, Shanghai Normal University, Shanghai, China. Her current research interests include social network analysis, trust, and reputation computation.



**Yanmei Gu** is currently pursuing her M.Sc. degree in the Department of Computer Science and Technology at Shanghai Normal University, China. Her current research interests include social networks and educational data mining.

**Qianqian Zhai** received her M.Sc. degree in computer science and technology from the Shanghai Normal University in 2021. She is currently an engineer in the Information Technology Division of HZ Bank, Co., Ltd., Hangzhou, China.

**Xiaoyang Guo** is currently an engineer in the Shanghai Newtouch Software Co., Ltd., Shanghai, China.

**Zhonghua Zheng** is currently the CEO of the Boryou Technology Co., Ltd., Shanghai, China.

**Chang Guo** received her B.Sc. degree in communication engineering from the Donghua University in 2014 and her Ph.D. degree in information and communication intelligent system from the Donghua University in 2020. She is currently a Lecturer of the Shanghai Normal University, Shanghai, China. Her research interests focus on solving the traffic prediction based on vehicular ad hoc networks, traffic load balance in road network, recommendation system and intelligent education.