

# Overlapping Community Detection Algorithm Based on the Law of Universal Gravitation

Chunde Yang

*School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, Sichuan, China*  
*Institute of System Theory and Application, Chongqing University of Posts and Telecommunications, Chongqing, Sichuan, China*

Mingjiang Li & Yongchao Wang

*School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, Sichuan, China*

**ABSTRACT:** Community mining has been the focus of many recent researches on complex networks. In this paper, the authors proposed a community mining algorithm based on Universal Gravitation principle (UGPCA). This algorithm based on the Law of Universal Gravitation is used to add the quality into the information of the nodes, and measure the weights of nodes in the networks. Each node will generate a gravitational field and have a force to the other nodes which can be reached. At last, the authors tested the UGPCA on real-world networks. Experimental results show the UGPCA has the ability of community mining and also has the ability to find the overlapping nodes.

**Keywords:** complex networks; community detection; Law of Universal Gravitation; overlapping node

## 1 INTRODUCTION

Complex networks have the properties of self-organization, self-similar, small-world, scale-free and so on. In the real world, many complex systems could be represented as networks, such as social networks, e-mail networks, protein interaction networks, scientific co-authorship networks[1]. These complex systems belong to different academic sectors; however, they both have the common properties of complex network.

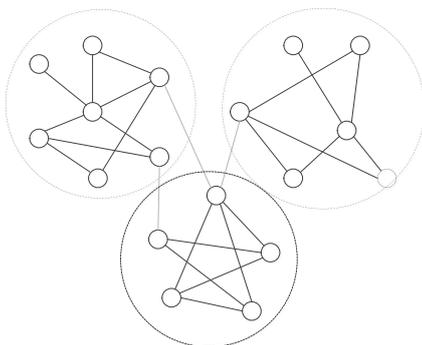


Figure 1. Community structure in the network

In the year of 2002, Girvan [2] and Newman proposed the community structure of social networks in

their paper. The research of community structure had attracted many scholars' attention. The community structure is a kind of topological property widely existed in complex network. The nodes in networks are always found to cluster into tightly-knit groups with a high density of within-group edges and a lower density of between-group edges. Figure 1 shows the community structure in a network. There are three communities in the figure divided by the dashed line.

## 2 RELATED WORKS

Community detection is a core problem in the research of community structure. In the real world, researchers widely applied the community detection to many areas, such as friends-recommend of social networks, search engine, prediction of disease transmission. The community detection algorithm has attracted a lot of attention in recent years. Chauhan et al [3] proposed a method of community detection by using the spectrum of network adjacency matrix. Evan et al [4] transformed edges into nodes, proposed the concept of edge graph, and detected community by reconstructing network. According to the probability theory, Ball et al [5] proposed a method based on mixed probability model. Because of the similarity between nodes in network and particles in physics, many scholars began to search the method of community detection from network dynamics. Inspired from the idea of data

fields, a community discovery algorithm based on topological potential is proposed in [6]. Shi et al [7] proposed a method based on potential function of data field and improved the robustness of Label Propagation Algorithm. Cheng et al [8] researched the relationship between diffusion dynamics on network and community structure and proposed a method to detect community structure by using network conductivity. From the point of Brownian motion, Zhou et al [9] formulated the measurement of similarity between nodes and proposed the algorithm of community detection. Based on the seed expansion and the Newtonian gravitation, Liu et al [10] proposed an overlapping community detection algorithm.

The existing algorithms based on network dynamics have good performance to detect community, whereas, they are not suitable for different networks. In real-world networks, nodes with hop 2 always have the certain influence on the current node. Based on the foundation of Law of Universal Gravitation, we take the influence of nodes with different distance into consideration and propose a community detection algorithm UGPCA. We compare the UGPCA with LMF algorithm, and ensure the algorithm is reasonable though experiment.

### 3 MODEL

Most networks could be transformed into mathematical symbols. A Undirected network could be represented by  $G=(V,E)$ , where  $V=\{v_1, \dots, v_n\}$  is the nonempty and finite set of nodes that exist in network  $G$ ,  $|V|=n$ ;  $E \in V \times V$  is the set of edges,  $|E|=m$ . In this part, we introduce the law of universal gravitation into network, and build the model of community detection.

#### 3.1 Law of Universal Gravitation

Law of Universal Gravitation is an important principle in classical physics, the formula is shown as follows:

$$F=G \frac{m_1 m_2}{r^2} \quad (1)$$

Its principle is that: the gravitation action exists between two particles, the size of gravitation is proportional to their mass product, and it is inversely proportional to their distance.

In the real-world network, the property of nodes always have many meanings, such as the influence of the nodes in social networks, the forward capability of the nodes in communication system, and the flow of the nodes in transportation network. Enlightenment by Law of Universal Gravitation, nodes in the network are regarded as particles, and add mass into the property of node. Thus, each node would have acting force with other reachable nodes.

#### 3.2 Node degree

Identifying the key spreaders in complex networks and

having the maximum impact on information dissemination is of significant impact on our understanding and control of spreading on networks. Node degree is a representative of local property of networks [11, 12].

On the graphic, node degree expresses a number of edges which are connected with current node. The areas always have dense edges where contain the nodes with large degree. Node degree could reflect the importance of nodes in the network. In the calculation method, the calculation of node degree also has a lower time complexity. As a result, we use node degree to describe the mass of node. We can obtain the value of node degree by calculating the adjacency matrix of network. The calculation formula of node degree is shown as follows:

$$k_u = \sum_{v \in V_u} w_{uv} \quad (2)$$

Where,  $w_{uv} = \begin{cases} 1, & u \text{ joined by } v \\ 0, & \text{others} \end{cases}$ ,  $V_u$  is the set of nodes

which adjacent to u.

We accomplish the cluster of nodes in the network by the gravitation among nodes. The mass of nodes are expressed as node degree, thus, the formula of force between node u and its adjacent nodes is shown as follows:

$$f(u) = G \sum_{v \in V_u} \frac{k_u k_v}{d^2(u, v)} \quad (3)$$

#### 3.3 Impact factor $\mu$

For computer networks, the distance between two nodes could not be described by Euclidean distance, but be described by hop count. As for node u, the hop count to its adjacent nodes is 1. The size of gravitation is related to the distance of nodes. We introduce the impact factor  $\mu$  to measure the influence of force among nodes. In the network, the sum of gravitation between node u and its adjacent nodes is shown as follows:

$$f(u) = \mu \sum_{v \in V_u} k_u k_v \quad (4)$$

In the real-world networks, nodes with hop 2 always have a certain influence on the current node. However, nodes with hop 3 always have rare influence. We consider the influence of nodes with hop 2 into algorithm, set  $\mu_1$  which is the impact factor of nodes with hop 1, and  $\mu_2$  which is the impact factor of nodes with hop 2. The sum of gravitation to the current node is shown as follows:

$$f(u) = \mu_1 \sum_{v \in V_u} k_u k_v + \mu_2 \sum_{w \in V_{u'}} k_u k_w \quad (5)$$

Where,  $V_u$  is the set of nodes that are adjacent with node u, and  $V_{u'}$  is the set of nodes with hop 2 to node u.

We apply the model to the community detection

algorithm of complex network. Based on the description mentioned earlier, the nodes with degree are greater than 0, and they will have gravitation. Formula (5) is used to calculate the acting force to the current node which is generated by the surrounded nodes. By analyzing the relationship of interaction among nodes, we obtain the community structure and algorithm convergence.

#### 4 ALGORITHM

We propose an improved algorithm UGPCA based on the Law of Universal Gravitation. In the algorithm, we consider the influence of nodes with hop 1 and hop 2, and separately set the different impact factor  $\mu_1$  and  $\mu_2$ . In the process of algorithm implementation, we introduce the sets  $V, C_1$  and  $V_r$ .  $V$  is the set contains all nodes in network,  $C_1$  is the set of nodes which have joined a community, and  $V_r$  is the set of remaining nodes.

Input: dataset network  $G=(V, E)$ , the impact factor  $\mu_1$  and  $\mu_2$  of nodes with hop 1 and hop 2, the coefficient  $\varphi$  of overlapping nodes. Where,  $V = \{v_1, \dots, v_n\}$ ,  $|E| = m$ .

Output: the number  $t$  of communities, communities  $C_1, \dots, C_t$ .

Process of algorithm:

1. Calculate the nodes degree in network  $V$ , and select the node  $u$  with largest degree as core node. If there are some same largest nodes degrees, we will compare the sum of degree of their adjacent nodes.

2. Divide the core node with its adjacent nodes into the community as an initial community. Add the nodes that joined the initial community into set  $C_1$ , and add the other nodes into set  $V_r$ .

3. Select the node  $w$  from  $V_r$  which is adjacent to the nodes in  $C_1$ . Calculate the sum of gravitation  $f(u)$  between  $u$  and nodes from  $C_1$ , and calculate the sum of gravitation  $f(u)'$  between  $u$  and nodes from  $V_r$ . If  $f(u) > f(u)'$ , we will add the node  $w$  to set  $C_1$ , and others remain the node  $w$  in set  $V_r$ . Set a

constant  $\varphi$ , in which,  $0 < \varphi < 1$ . If  $\frac{\min(f(u), f(u)')}{\max(f(u), f(u)')} > \varphi$ ,

we will determine the node which is an overlapping node.

4. Repeat the process 3, traversal the nodes of set  $V_r$ . If there are no nodes meet the conditions to join the initial community, we will complete a community partition. If  $|V_r| \neq 0$ , we will repeat the process mentioned earlier until  $|V_r| = 0$ .

## 5 EXPERIMENTS AND RESULTS

Experimented with the algorithm in the real-world networks and compared the effect of UGPCA with algorithm of LMF by using modularity, it is proven that the UGPCA is feasible and effective. In the experiment, we will use the dataset of American College Football network, Zachary Karate Club network and Dolphin network, and we separately test the effectiveness of the chosen impact factor, the division effect and so on.

### 5.1 Selection and discussion of parameters in algorithm

We will discuss the impact factor  $\mu_1$  and  $\mu_2$  in this experiment. First, we select different parameters in the experiment. Then, we compare the division effect of American College Football network through the parameters, and obtain the impact factor with best division effect.

In the formula (5), we set  $\mu_1 + \mu_2 = 1$ . Figure 2 shows the accuracy curve of node detection. From this curve, when  $\mu_2 = 0.1$ , the accuracy of node detection reaches the best one. With the increase of impact factor of nodes with hop 2, the accuracy rapidly decreases. Thus, we can draw the conclusion that the nodes with hop 2 have certain and weak influence on the network partition.

The division accuracy with different impact factor

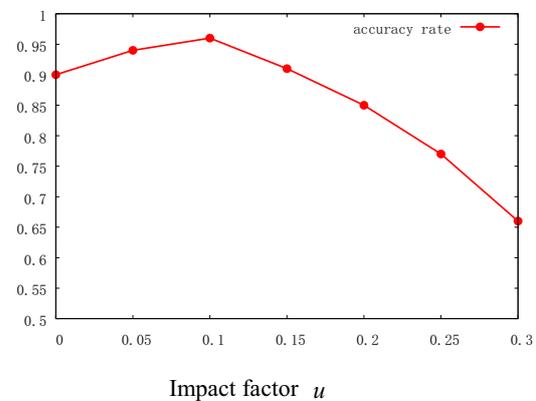
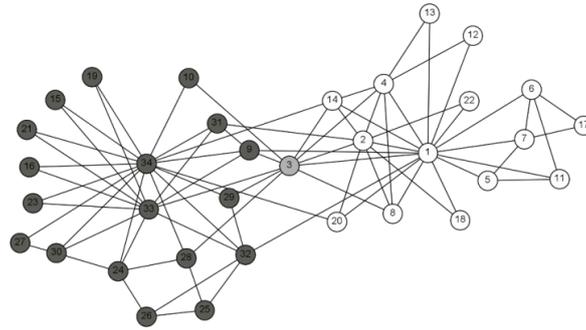


Figure 2. Accuracy rate with different impact factors

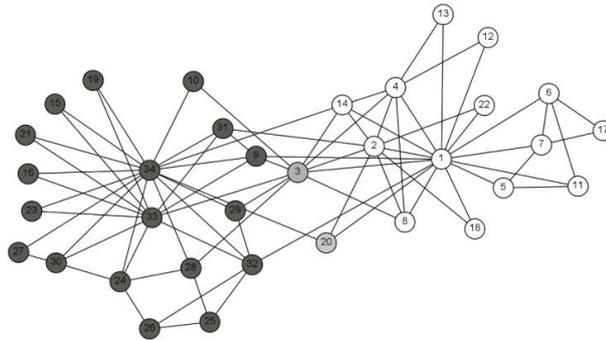
### 5.2 Partition result in Zachary dataset

Zachary karate club is one of the most commonly-used dataset in the research of community detection. This network contains 34 nodes and 78 edges. The network structure of karate club is shown in figure 3(a), of which, the light color nodes and the deep color nodes are respectively expressed two communities. Node 1 and node 34 are the core nodes of two communities; node 3 is the overlapping node.

We use UGPCA to divide the Zachary karate club network and select the parameter  $\mu_2 = 0.09$ ,  $\varphi = 0.93$ .



(a)Zachary Karate Club network structure



(b)Network structure detected by UGPCA

Figure 3. Using UGPCA to divide network

Figure 3(b) shows the result which is divided by UGPCA. The network are divided into two communities, node 3 and node 20 are affirmed to be overlapping nodes. The result demonstrates that UGPCA is effective and accurate.

### 5.3 Formula of force among nodes

Newman et al proposed the modularity method to evaluate the rationality of community detection. The foundation of modularity is that the expectation of edge number in the community is higher than the expectation in a random network. In order to solve the heterogeneity of nodes distribution, Newman et al [13] proposed an improved modularity method, the definition is shown as follows:

$$Q = \frac{1}{2m} \sum_{ij} \left( A_{ij} - \frac{k_i k_j}{2m} \right) \delta(C_i, C_j) \quad (6)$$

Where,  $A_{ij}$  is the element in the adjacency matrix

of network,  $k_i$  is the degree of node  $i$ ,  $C_i$  is the

community that contains node  $i$ ,

$$\delta(C_i, C_j) = \begin{cases} 1, & i \text{ and } j \text{ in the same community} \\ 0, & \text{others} \end{cases}$$

We compare the UGPCA with algorithm of LMF by using the modularity method in the formula (7). The dataset in the experiment is used in American College Football network, Zachary Karate Club network and Dolphin network. Set the parameters  $\mu_1=0.9$ ,  $\mu_2=0.1$  in UGPCA. Table 1 shows the comparison result between UGPCA and LMF by using modularity method.

From table 1, we can draw the conclusion: In the network of Zachary and Football, the Q value of UGPCA is higher than LMF; it shows that the networks divided by UGPCA are more significant. But in the network of Dolphin, the UGPCA is weaker than the LMF.

Table 1 Comparison between UGPCA and LMF

Algorithm	UGPCA	LMF
Data Set	Modularity Q	
Football	0.354	0.336

<b>Zachary</b>	0.442	0.423
<b>Dolphin</b>	0.298	0.323

- [13] Clauset A, Newman M E J. & Moore C. 2004. Finding community structure in very large networks. *Phys. Rev. E.*, 70(6): 066111.

## 6 SUMMARY

Community detection algorithm is an important point in the research of complex network. In the present work, most of algorithms based on network dynamics model are just considered as the influence of adjacent nodes. In this paper, based on the Law of Universal Gravitation, a new algorithm was proposed for detecting overlapping community. The algorithm is used to calculate the mass of nodes, and consider the influence of nodes with different hop count. On the problem of overlapping nodes, the algorithm has a standard to judge overlapping nodes. The result of experiment on real-world networks shows that the algorithm could effectively and precisely detect community.

## REFERENCES

- [1] Y Bo, L Da-You. & L Jiming, et al. 2009. Complex Network Clustering Algorithms. *Journal of Software*. 20(1): 54-66.
- [2] Girvan M. & Newman M E J. 2002. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA*. 99(12): 7821-7826.
- [3] Chauhan S, Girvan M. & Ott E. 2009. Spectral properties of networks with community structure. *Phys. Rev. E.*, 80(5): 2797-2804.
- [4] Evans T S. & Lambiotte R. 2009. Line graphs, link partitions, and overlapping communities. *Phys. Rev. E.*, 80(1): 106105.
- [5] Ball B, Karrer B. & Newman M E J. 2011. An efficient and principled method for detecting communities in networks. *Physical Review E.*, 84(3): 1-14.
- [6] G Wen-Yan, H E Nan. & L I De-Yi, et al. 2009. Community Discovery Method in Networks Based on Topological Potential: 20, 2241-2254.
- [7] S Lixin. & Z Junxing. 2014. Label propagation algorithm based on potential function for community detection. *Journal of Computer Applications*. 34(3): 738-741.
- [8] Cheng X Q. & Shen H W. 2010. Uncovering the community structure associated with the diffusion dynamics on networks. *J. Stat. Mech.*, P04024.
- [9] Zhou H J. 2003. Network landscape from a Brownian particle's perspective. *Phys. Rev. E.*, 67(4): 041908.
- [10] Q Liu. 2014. Overlapping community detection algorithm based on expansion of gravitational degree. 35(3): 852-856.
- [11] Liu Jian-Guo, Ren Zhou-Ming, Guo Qiang. & Wang Bing-Hong. 2013. Node importance ranking of complex networks. *Acta Physica Sinica*, 62(17): 178901.
- [12] Z Zhao, H Yu. & Z Zhu, et al. 2014. Identifying Influential Spreaders Based on Network Community Structure. 37(4): 753-766.