

Community Discovery Research Based on Louvain Algorithm

Xiaoqian You^{1,a}, Banghui Yin^{1,b}

¹College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 500000, China;

^ayouxq@cqupt.edu.cn, ^bky15282101730@163.com

Keywords: community structure, Louvain, community discovery.

Abstract: Aiming at the problems of low community quality and low efficiency of the existing community discovery algorithms, this paper improves the community discovery algorithm by selecting the seed nodes based on the Louvain algorithm. Experimental results show that the improved community discovery algorithm has greatly improved the quality of community and the efficiency of the algorithm.

1. Introduction

If the entities in complex systems are taken as nodes, and the relations between entities are taken as edges, a network that can represent the internal relations of complex systems can be constructed, which is called complex networks. Community structure is one of the most common topological structures in complex networks. According to the characteristics of community structure, the community structure in complex network is mined by some methods. This process is called community discovery. Community discovery algorithm is called community detection algorithm. Through the study of community discovery, it is of great significance to deeply understand the network topology, access to network information and predict network behavior.

Community discovery algorithm is mainly divided into Graph Segmentation Algorithm and hierarchical clustering algorithm. Hierarchical clustering algorithms are divided into 2 categories: splitting algorithm and agglomerative algorithm. Typical agglomerative algorithms include FN (fast Newman) algorithm[1] and Louvain algorithm[2], in which Louvain algorithm is considered as one of the best performance community discovery algorithms. Wu Zufeng et al [3] proposed an improved Louvain community partitioning algorithm to reduce the running time of the algorithm by pruning the nodes. Zhao Runqian and other[4] methods optimize the efficiency of community discovery by combining the method of seed node pre selection.

In this paper, based on the Louvain algorithm, the scale-free and small world characteristics combined with the network, in the implementation process to choose strong nodes as seed nodes, to prevent a large community over the merger, at the same time to merge small communities, while reducing the number of iterations.

2. Louvain Algorithm

Community discovery is a complex and meaningful process, which plays an important role in studying the characteristics of complex networks. In recent years, many community scholars have paid close attention to discovering and analyzing community structures in complex networks, and many community discovery algorithms have emerged.

Figure $G = G(V, E)$, so called community discovery, refers to the determination of $nc(\geq 1)$ community in figure G

$$C = \{C_1, C_2, \dots, C_{nc}\} \quad (1)$$

The vertex set of each community forms a cover of V . If the intersection of the vertex sets of any two communities is empty, it is called C non overlapping community; otherwise, it is called overlapping community.

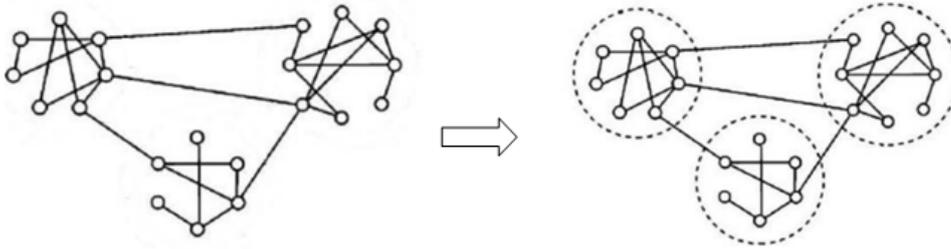


Fig. 1 Mathematical description of community discovery

The modularity of community network [5] is a measure for evaluating the quality of a community network, and its range of value is (0,1), which is defined as follows:

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

In the upper form, A_{ij} represents the edge weight between the nodes i and j (when the graph does not have the right, the edge weight can be regarded as 1). k_i represents the edge weight of adjacent edge of node i and (when graph is not right) is the degree of node. m is the edge weight of all edges in a graph. C_i is the community number where the node i is located. $\delta(C_i, C_j)$ indicates that the node i and j are in the same cluster, and when $C_i = C_j$, $\delta(C_i, C_j) = 1$. The closer the Q value to 1, the higher the quality of the community.

Each node in the community network allocates a community number. At the initial stage, each node is regarded as a community, corresponding to the modular increment ΔQ [6] of the community, which is defined as follows:

$$\Delta Q = \left[\frac{\sum in + K_{i,in}}{2m} - \left(\frac{\sum out + k_i}{2m} \right)^2 \right] - \left[\frac{\sum in}{2m} - \left(\frac{\sum out}{2m} \right)^2 - \left(\frac{k_i}{2m} \right)^2 \right] \quad (3)$$

In the formula, $\sum in$ is the organization's internal edge weight and node; $K_{i,in}$ is i and the community is connected to the C edge weights and $\sum out$; all nodes and associations connected to the internal edge weights and the weights of all k_i ; and is connected with the i side.

The Louvain Algorithm has 2 Stages: Phase One, Each node in the graph is considered as an independent community, and the number of communities is the same as the number of nodes. Each node of the i , in turn try to node i assigned to each of its neighbor nodes in the community, calculate the distribution before and after allocation of incremental ΔQ module, and record the neighbor node, the largest q if $\Delta Q > 0$, put the neighbor node distribution of i ΔQ 's largest community, otherwise unchanged; repeat the operation until all nodes belong to the community no longer change. Phase Two, The compression of the map, all the nodes in the same community compressed into a new node, the weights of nodes in the same community of edges into the weight of the new node ring, edge weights between communities into the new edge weights between nodes; repeat 1 until the entire graph module of no change.

3. Improved Louvain Algorithm

Louvain complex network has scale-free and small world effect, the community structure characteristics. The scale-free property refers to the social network node degree showed a power-law distribution, a high degree of nodes in network, and most of the nodes in a low degree. In complex

networks, high degree nodes tend to can attract a low degree of adjacent nodes, so as to form a community.

The improved community discovery algorithm has 2 stages, phase 2 is the same as phase 2 of Louvain algorithm, and its improvement is mainly concentrated in stage 1. In the improved algorithm, stage 1 contains 2 steps. Firstly, the seed nodes are selected

$$G = \{v | \text{deg}(v) > p + q\} \quad (4)$$

$\text{deg}(v)$ denotes the degree of node v ; p denotes the average degree of node; q denotes the standard variance of the degree of node. Then each node each into a community, for each non seed node i to node i , tried in turn assigned to the community where the seeds, calculate the distribution before and after allocation of incremental ΔQ module, and record the seed node, the largest ΔQ if $\Delta Q > 0$, put the seed node neighborhood node i assignment ΔQ is the largest, or remain unchanged.

The Improved Louvain Algorithm has 2 Stages: Phase One, According to formula (4), the seed node set is obtained. Then each node each into a community, for each non seed node i to node i , tried in turn assigned to the community where the seeds, calculate the distribution before and after allocation of incremental ΔQ module, and record the seed node, the largest ΔQ if $\Delta Q > 0$, put the seed node neighborhood node i assignment ΔQ is the largest, or remain unchanged. Phase Two, The compression of the map, all the nodes in the same community compressed into a new node, the weights of nodes in the same community of edges into the weight of the new node ring, edge weights between communities into the new edge weights between nodes; repeat 1 until the entire graph module of no change.

The time complexity of Louvain algorithm is approximately $O(|V| \ln |V|)$. Because the algorithm is optimized on the basis of Louvain algorithm, the algorithm runs faster than the Louvain algorithm.

4. Experiment and Result Analysis

4.1 Experimental Scheme

In order to evaluate the improved community detection algorithm, this paper selects 6 sets of real complex network data sets and Zhao Runqian and other improved algorithms and Louvain algorithm for experimental comparison. The specific information of the data set is shown in table 1. The experiment analyzes and compares the algorithm from the community division effect and efficiency, and uses 3 evaluation indexes: modularity value Q , community number and algorithm running time.

Table 1 Community discovery experimental data set

Data Set	Facebook1	Facebook2	Hamsterster	Advogato	Youtube	Google+
nodes	48	1893	2000	5042	13679	23613
edges	1962	3835	16098	39227	76741	39182

4.2 Experimental results and analysis

In accordance with the above experimental scheme, the improved algorithm and Zhao Runqian algorithm, Louvain algorithm experiment results are shown in table 2. The algorithm evaluation indexes include Q , community number and algorithm running time.

Table 2 Comparison of experimental results of algorithms

Evaluating Indicator	Data Set	Algorithm in this paper	Zhao Runqian	Louvain Algorithm
Q	Hamsterster	0.5326	0.5248	0.5079
	Facebook1	0.5645	0.5432	0.5399
	Facebook2	0.3566	0.3333	0.3205
	Advogato	0.4125	0.3926	0.3829
	Youtube	0.6958	0.6767	0.6440
	Google+	0.7356	0.7282	0.6459
nodes	Hamsterster	28	30	131
	Facebook1	4	4	7
	Facebook2	12	14	7
	Advogato	18	20	196
	Youtube	27	29	604
	Google+	40	41	101
Function Time/s	Hamsterster	1.96	2.02	2.53
	Facebook1	0.08	0.09	0.11
	Facebook2	2.51	2.53	3.38
	Advogato	18.25	18.52	21.99
	Youtube	128.69	135.14	267.59
	Google+	158.67	168.71	212.26

Can be seen from table 2, the improved communitydiscovery algorithm Q value than Zhao Runqian algorithm and Louvain algorithm, the number of community than Zhao Runqian algorithm and Louvain algorithm, the efficiency is better than Zhao Runqian's improved algorithm and Louvain algorithm.

5. Conclusions

The existing community discovery algorithms have many iterations, slow running time, unreasonable community division, excessive merger of large communities, and no timely merging of small communities. This paper improves the community discovery algorithm based on Louvain algorithm, reduces the running time of the algorithm, and improves the quality of community division.

References

- [1] Newman M E J. Fast algorithm for detecting community structure in networks[J]. Physical Review E Statistical Nonlinear & Soft Matter Physics, 2004, 69(6): Article No.066133.
- [2] Blondel V D, Guillaume J L, Lambiotte R, et al. Fast unfoldingof communities in large networks[J]. Journal of Statistical Mechanics Theory & Experiment, 2008, 2008(10): 155-168.
- [3] Wu Zufeng, Wang Pengfei, Qin Zhiguang, et al. Improved algorithm of Louvain communities dipartition[J]. Journal of University of Electronic Science and Technology of China, 2013,42(1): 105-108.
- [4] Zhao Runqian, Wu Yu, Chen Xin. Community discovery and visualization algorithms for large scale social networks [J]. Journal of computer aided design and computer graphics, 2017, 2 (29):328-336.
- [5] Girvan M, Newman M E J.Community structure in social and biological networks[J]. Proceedings of the National Academy of Sciences,2002,99(12) : 7821-7826.
- [6] Palla G, DeRenyi I, FaRkas I, et al. Uncovering the overlapping community structure of complex networks in nature and society[J]. Nature,2005,435(6) : 814-818.