

A Community Detection Algorithm Based on Local Double Rings and Fireworks Algorithm

TianRen Ma and Zhengyou Xia^(✉)

College of Computer Science and Technology,
Nanjing University of Aeronautics and Astronautics, Nanjing 210015, China
zhengyou_xia@nuaa.edu.cn

Abstract. In recent years, more and more algorithms have been proposed to detect communities. An improved community detection algorithm based on the concept of local double rings and the framework of fireworks algorithm (LDRFA) has been proposed in this paper. Inspired by the framework of FWA, an improved distinctive fireworks initialization strategy was given. We use this strategy to obtain a more accurate initial solution. Secondly, on the basis of fireworks algorithm, the amplitude of explosion was used to calculate the probability of changing node label. Thirdly, the mutation operator was proposed. Nodes chose labels based on the idea of LPA. Finally, tests on real-world and synthetic networks were given. The experimental results show that the proposed algorithm has better performance than existing methods in finding community structure.

Keywords: Community detection · Fireworks algorithm · Swarm intelligence

1 Introduction

Complex networks have attracted increasing attention of researchers from different fields [1]. Community detection problem is an NP-hard one [2] and has become a research hotspot in recent years.

Label propagation algorithm is one of the most widely used community partitioning algorithms. The most primitive label propagation algorithm [3], proposed by Raghavan et al., has a high accuracy rate and it runs very fast. Swarm intelligence optimization algorithms have great performance on solving lots of real world problems and attract many scholar's attentions. Various algorithms for community detection were proposed by using swarm intelligence optimization algorithms. GA-Net is one of the first approaches to solve detection problem which was proposed by Pizzuti [4]. GA-Net is a single-objective algorithm and optimizes a simple fitness function to identify densely connected groups. CNM was proposed by Clauset et al. in 2004 [5]. This algorithm is a fast greedy modularity optimization algorithm and a fast implementation of the original Girvan-Newman algorithm. Infomap is introduced by Rosvall and Bergstrom [6]. This algorithm used a new information theoretic method to find community structure in a complex network. The Fireworks Algorithm (FWA) [7] is one of the latest swarm intelligence optimization algorithms which was proposed by Tan in 2010. It is a metaheuristic

swarm intelligence algorithm and simulates the explosion of fireworks and the generation of the sparks.

The remainder of this paper is organized as follows. The framework of fireworks algorithm is described in Sect. 2. In Sect. 3, LDRFA is particularly studied, followed by experiments in Sect. 4. Finally, we conclude the paper in Sect. 5.

2 The Framework of Fireworks Algorithm

The fireworks algorithm (FA) was inspired by the behavior of firework. The explosion process of firework can be regarded as the process of searching in a local space around a specific point. When a firework explodes, lots of sparks are generated around it. The number and amplitude of generated sparks will be designed. For each generation of explosion, n locations for n fireworks were selected. After explosion, every firework has generated sparks around it already. The number and amplitude of sparks are designed by the explosion operator. Finally, when the optimal location is found, the algorithm stops. Otherwise, the algorithm keeps the best fireworks or sparks and selects the rest $(n-1)$ fireworks or sparks for the next generation according to their distances. The framework of FWA can be divided into four parts: the explosion operator, the mapping rule, the Gaussian explosion (mutation) operator and the selection strategy. The explosion operator was described as follow.

2.1 Explosion Operator

At each iteration, each firework generates sparks using this explosion operator. This explosion operator plays a key role in the whole algorithm because that it is responsible for calculating the number and amplitude of sparks.

The number of sparks S_i is calculated using the following formula:

$$S_i = m \times \frac{y_{worst} - f(x_i) + \varepsilon}{\sum_{i=1}^n (y_{worst} - f(x_i)) + \varepsilon} \quad (1)$$

Where m is a constant to control the total number of sparks and y_{worst} is the worst value of the objective function among the n fireworks. Function $f(x_i)$ represents the fitness value of individual x_i . The last parameter ε is used to prevent the denominator from becoming 0.

The amplitude of sparks A_i is calculated as follow:

$$A_i = A' \times \frac{f(x_i) - y_{best} + \varepsilon}{\sum_{i=1}^n (f(x_i) - y_{best}) + \varepsilon} \quad (2)$$

Where A' denotes the sum of all amplitudes, while y_{best} is the best value of the objective function among the n fireworks. The meaning of $f(x_i)$ and ε are the same as mentioned in formula (1).

3 LDRFA: A Community Detection Algorithm Based on Local Double Rings and Fireworks Algorithm

In this paper, the number of positions is equal to the number of nodes in the network. An arbitrary fireworks x_i represents the community detection result of the network and x_i^k represents the label of node k .

3.1 Concept of Node Importance

Inspired by the PageRank algorithm [8], a new method to measure the importance of nodes is designed.

Definition 1. (Node Importance). Given a network $G = \{V, E\}$, for any node $v_i \in V$, the node importance is:

$$NI(v_i) = \sum_{v'_i \in V'} \frac{1}{d(v'_i)} \quad (3)$$

V'_i is the neighbor node set of node v_i and $d(v'_i)$ is the degree of node v'_i . The importance of a node v_i depends on the number and degree of its neighbor nodes. If the degree of the neighbor node is larger, the contribution to the importance of the node v_i is relatively smaller and vice versa.

3.2 An Improved Fireworks Initialization Method

In LDRFA, we propose an improved fireworks initialization method. For convenience of description, we denote n as the total number of fireworks, V' as the sorted set of nodes and $V'[i]$ as the i th node in V' .

First of all, the total number of fireworks n is equal to 50. After the whole nodes were sorted in descending order by their importance and initializing the label of node, for each firework x_p , the first local ring is the shortest ring containing $V'[x_i.id \bmod 3]$. The second ring is the shortest ring containing the most important node which is not $V'[x_i.id \bmod 3]$ in the first ring. If there are more than one rings available, then randomly choosing one. After the local double rings were both found, we change the label of nodes in these two rings and the label of the most important neighbor of them to the label of $V'[x_i.id \bmod 3]$. Then the number and amplitude of sparks were calculated according to the formulas (1) and (2). Then the fireworks initialization process is over.

3.3 Detailed Description of LDRFA

LDRFA is described as follows:

(1) Initializing the label of each node in the network; calculating their importance and sorting them in descending order by importance. Setting $m = 50$, $A' = 40$, the total number of fireworks $n = 50$. Setting the maximum number of iterations to 100.

(2) Initializing the fireworks according to Sect. 3.2.

(3) Generating sparks and selecting fireworks or sparks for next generation. This part does not stop until the number of iterations reaches its maximum.

In LDRFA, the amplitude of fireworks is used to control the probability of changing node label. And the node label updating strategy updates node's label based on its neighborhood labels. The label with maximum frequency is used as new label. We denote this label as Neig_label_k . x_i^k in this part represents the label of node k in firework i .

$$x_i^k = \begin{cases} \text{Neig_label}_k, & \text{if } \text{random}(0, 1) < \text{sigmoid}(A_i) \\ x_i^k, & \text{else} \end{cases} \quad (4)$$

$\text{random}(0,1)$ is a randomly number between 0 and 1. $\text{sigmoid}(A_i)$ is the sigmoid function of the amplitude A_i .

After all of sparks were generated, the top n fireworks and sparks which were sorted in descending order by their modularity value will be selected into next iteration. Finally, when the termination condition was reached, the community structure represented by the fireworks or sparks with maximum modularity value was accepted.

4 Experiments Result and Analysis

In this section, the performance of LDRFA is tested on standard data sets which are commonly used in the real world and the synthetic benchmark networks. The former community data is collected from the real world and often measured by modularity [9]. The latter community data is a data set that are constructed from pre-set parameters. The accuracy of this community detection result can be measured by NMI.

4.1 Experimental Result of the Real World

There are many real-world networks that have been abstracted into community detection fields, such as Zachary's Karate Club, which is constructed by observing an American university karate club. In addition to the above network, we also choose the Dolphin social network. It is the dolphin social network that D. Lusseau et al. used for seven years to observe the dolphin population of Doubtful Sound, New Zealand. The third network we choose is Polbooks, which was created by V. Krebs. In this section, we use these three networks to perform community partitioning experiments. The detail parameters of these networks are shown in Table 1.

We use four different algorithms for community detection on these three networks. The other three algorithms are GA-Net, CNM and Infomap which are mentioned in Sect. 1. The values of modularity obtained from the experiment are shown in the following table (Table 2).

Table 1. Real-world networks

Name	Nodes	Edges	Average Degree
Karate	34	78	4.59
Dolphins	62	159	5.13
Polbooks	105	441	8.40

Table 2. Comparison of modularity values

Networks	Q(LDRFA)	Q(GA-Net)	Q(CNM)	Q(Infomap)
Karate	0.3949	0.4019	0.3990	0.3989
Dolphins	0.5264	0.4701	0.4496	0.4310
Polbooks	0.5262	0.5798	0.4193	0.4198

The proposed LDRFA algorithm in the Dolphins network has achieved the highest values of modularity and in the Polbooks network also achieved a nice performance.

4.2 Experimental Result of the Synthetic Benchmark Networks

In this section, we compared the performance on the GN extended benchmark networks. GN extended benchmark networks proposed by Lancichinetti et al. [10] is an improved version of the classical GN benchmark networks. There is an important mixing parameter μ , which control the tightness of connections between different communities.

We can find that these four algorithms have perfect performance when $\mu \leq 0.15$. With the increase of μ , Infomap algorithm first became unsuccessful and can't discover community structure when μ is greater than 0.25. The detection ability of GA-Net and CNM algorithms began descending when $\mu > 0.3$ and $\mu > 0.2$ respectively. But LDRFA

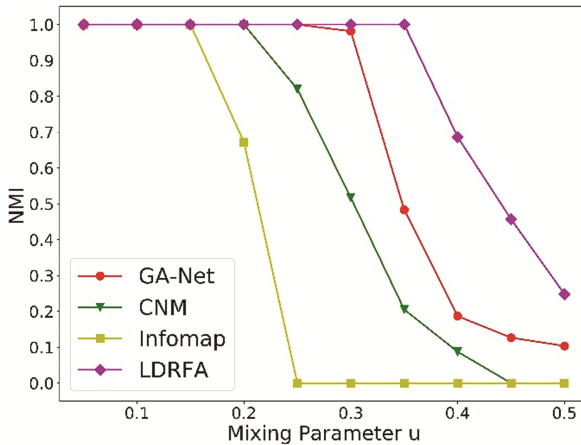


Fig. 1. Average NMI values on GN extended networks

always had a perfect performance when $\mu \leq 0.35$ and a good detection result was obtained when $\mu = 0.4$ (Fig. 1).

The average modularity values obtained by these four algorithms are shown in Fig. 2. From the figure we can see that all algorithms get a consistent modularity value when $\mu \leq 0.15$. As μ increases, the detection task becomes more and more difficult. When $\mu = 0.2$, the Infomap algorithm gave a worst modularity value of 0.32, but others gave a best value of 0.55. When μ is not less than 0.25, the performance of the other three algorithms is significantly reduced. Infomap and CNM algorithms fail to detect the community structure when $\mu = 0.35$. At the same time, the proposed algorithm still detected the community structure successfully with modularity value of 0.4013. When μ continues to increase, the performance of these algorithms began to decline rapidly because the community structure became ambiguous. From these two figures, the advantage of LDRFA against others was shown clearly.

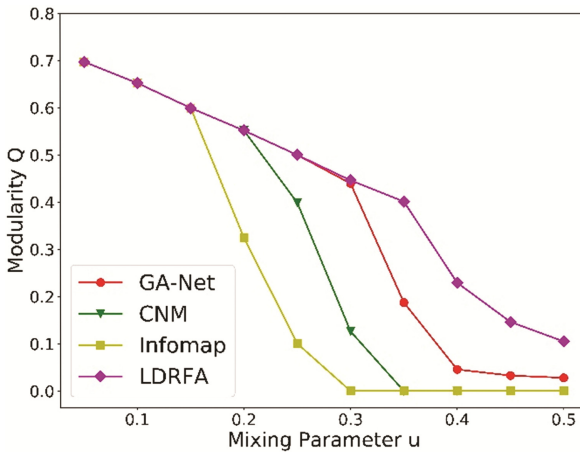


Fig. 2. Average modularity values on GN extended networks

5 Summary

In this paper, a community detection algorithm based on local double rings and fireworks algorithm is proposed to solve the problem of disjoint community partitioning, and we compare LDRFA with GA-Net, Infomap and CNM by using real-world networks and the synthetic networks.

The fireworks algorithm is a new swarm intelligence algorithm which has been widely applied to many fields. The LDRFA improves the initialization operation of fireworks in FA and proposes a new and effective method to discover small groups based on the concept of local double rings. In our algorithm, nodes are sorted by computing the importance of them and the initialization process proposed in this paper can improve the accuracy of detection result. On the basis of fireworks algorithm, the amplitude of explosion was used to calculate the probability of changing node label based on the idea

of LPA. Finally, the firework with highest modularity value was accepted and nodes with the same label were considered in the same community.

The experiment in this paper is divided into two parts. The value of modularity and NMI are used to evaluate the performance. Experimental results of these networks show that the algorithm proposed in this paper has achieved better performance.

References

1. John, H., et al.: Natural communities in large linked networks. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 541–546. ACM (2003)
2. Fortunato, S.: Community detection in graphs. *Phys. Rep.* **486**(3–5), 75–174 (2010)
3. Raghavan, U.N., Albert, R., Kumara, S.: Near linear time algorithm to detect community structures in large-scale networks. *Phys. Rev. E Stat. Nonlin Soft Matter Phys.* **76**(2), 036106 (2007)
4. Pizzuti, C.: GA-Net: A Genetic Algorithm for Community Detection in Social Networks. In: Rudolph, G., Jansen, T., Beume, N., Lucas, S., Poloni, C. (eds.) PPSN 2008. LNCS, vol. 5199, pp. 1081–1090. Springer, Heidelberg (2008). doi:[10.1007/978-3-540-87700-4_107](https://doi.org/10.1007/978-3-540-87700-4_107)
5. Clauset, A., Newman, M.E., Moore, C.: Finding community structure in very large networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **70**, 2 (2004).066111
6. Rosvall, M., Bergstrom, C.T.: Maps of random walks on complex networks reveal community structure. *Proc. Nat. Acad. Sci. U.S.A* **105**(4), 1118–1123 (2007)
7. Tan, Y., Zhu, Y.: Fireworks Algorithm for Optimization. In: Tan, Y., Shi, Y., Tan, K.C. (eds.) ICSI 2010. LNCS, vol. 6145, pp. 355–364. Springer, Heidelberg (2010). doi:[10.1007/978-3-642-13495-1_44](https://doi.org/10.1007/978-3-642-13495-1_44)
8. Brin, S., Page, L.: Reprint of: The anatomy of a large-scale hypertextual web search engine. *Comput. Netw.* **56**(18), 3825–3833 (2012)
9. Newman, M.E.J.: Fast algorithm for detecting community structure in networks. *Phys. Rev. E: Stat., Nonlin, Soft Matter Phys.* **69**(6), 066133 (2004)
10. Lancichinetti, A., Fortunato, S., Radicchi, F.: Benchmark graphs for testing community detection algorithms. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **78**(2), 046110 (2008)