

Complex Network Community Mining based on Genetic Algorithm

Chang Hao

Tai Yuan University, Department of Computer Engineering, Shanxi, China
changhao68@yeah.net

Abstract

The complex network community mining is a multidisciplinary research hotspot. It has been widely used in a terrorist organization to identify, protein function prediction, the metabolic pathways forecast, web community mining and link prediction in many areas. This article is mainly based on the genetic algorithm of the Community mining research oriented edge gene bit fast and effective local search variation algorithm. The algorithm uses a graph-based encoding strategy, the modules function as the objective function, Markov random walk method has certain the community divided accuracy and diversity of the initial population. The experiments show that the algorithm of the search space is effectively reduced, so the search efficiency can be further raised.

Keywords: Complex Network; Community Mining; Genetic Algorithm

1. Introduction

Many complex systems in the real world exists in the form of a complex network, or can be converted for the processing of complex networks, such as social networks, biological networks, Web network technology network. Basic statistical properties of the complex network has attracted many researchers in different fields, and has become one of the most important field of multidisciplinary research complex network analysis [1-3].The small-world effect refers to the complex network which has the characteristics of short path length and high clustering coefficient. The scale-free property is a complex network node degree follows a power law distribution characteristics, and community structure characteristics refers to the complex network generally has the closely interconnected nodes within the same community, the characteristics of the junction between the different communities interconnected sparse. With the application of the different areas of the community structure has different connotations, such as community representatives in the social network crowd with some similar characteristics, functional groups in biological networks reveal biological organization modules with similar functionality, Web network the document class clusters contain a large number of Web documents related topics. Complex network community mining purpose is to detect and reveal the structure of the inherent heterogeneous complex network of Community research has important theoretical and practical significance of the problem. It is not only to attract a large number of researchers from different disciplines, but also has been applied terrorist organizations to identify, protein function prediction, the metabolic pathways forecast, the Web community mining and link prediction in many areas[4-6].

2. Genetic Algorithm based on Local Search

The 2004 Newman used to characterize network community structure of the merits of a metric, called modularity function Q [7]. Since then, the combinatorial optimization

methods of the Q function as the objective function has become one of the mainstream methods of detecting network community structure, such as FN[8], SA[9], MSA[10], FUA[11] and LPAm[12] and so on. Maximize Q function is NP-complete, so the above methods are approximate optimization algorithm. The most recent years, as an effective method for solving NP-hard problem, genetic algorithm is gradually applied to the network community detection field. The current community detection algorithm is based on genetic algorithm using string encoding[13] and based on gene bit adjacent coding LAR (locus-based adjacency representation)[14, 15] in two ways. String encoding genetic algorithm is difficult to achieve crossover operation. LAR is well using the traditional cross LAR coded genetic algorithm operator, but it is difficult to give a fast and efficient mutation operator. For defects of genetic algorithm for solving network community detection problem, the article is that as the objective function of Q function and using LAR encoding proposed a genetic algorithm based on local search GALS (Genetic Algorithm with Local Search). The algorithm is using Markov random walk method to generate the initial population, and then executed by iteration uniform crossover, local search variation and $\mu + \lambda$ three genetic operators to detect network community structures. For the current existing variation method is inadequate. Firstly, this article gives the concept of edge chromosomal loci, in turn, based on in-depth analysis of the function Q proposed to an oriented edge gene bit fast and effective local search variation method. In addition, the GALS genetic strategy makes a spanning sub graph of each LAR the chromosome corresponding complex network population, so the algorithm of the search space is effectively reduced, so as to further improve its search efficiency.

2.1. Definition of Algorithm

Newman proposed for quantitative criteria to characterize network community structure of the merits. The quantitative criteria have been related for researchers in the field of widely accepted idea of modularity function Q comes from a phenomenon. The community structure of the network is more obvious, the greater the difference between the random network. It is defined as the actual connections in the network community within the community under the desired connection data the difference in the number of randomly connected. So the article mentioned genetic algorithm GALS, which is used Q function as the objective function and fitness function.

Given a undirected no right network $N(V, E)$, Point set V is assumed to be divided into a number of communities, If the network in any of the label of a node i to $r(i)$, the community to which it belongs as $c_{r(i)}$, Q function can be defined as:

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

Where $A = (A_{ij})_{n \times n}$ said the adjacency matrix of the network N , if the node i and the node j has side connection, then $A_{ij} = 1$, otherwise $A_{ij} = 0$; for the function $\delta(u, v)$, if $u = v$, its value is 1, otherwise the value is 0; k_i represents the degree of node i , defined $k_i = \sum_j A_{ij}$; $m = \frac{1}{2} \sum_{ij} A_{ij}$, represents the total number of edges in the network N .

2.2. Encoding

The GALS of the article is used to base adjacent chromosomal loci coding strategy LAR. The encoding is from the literature [17], which has been used for multi-objective clustering [18] and the detection of network community structure. For graph-based

encoding, each individual in the population contains n genes $(1, \dots, n)$, where n represent the number of nodes in the network, each gene bit i can take any value, $j \in \{1, \dots, n\}$. Each gene in the chromosome corresponds to a node in the network N , and if the value of the i gene in the chromosome is j , which can be interpreted as in the chromosome in the figure corresponding to the node i and j exists between the side links, That is, in the clustering result of the individual corresponding node i and j , will be located within the same community. Obviously, the pairs of chromosomes of the decoding process are to identify all the communicating branch in it corresponds to the figure, and located on the same node connectivity within the branch assigned to the same community. For [19] show that chromosome decoding process can be completed within a linear time. LAR coding strategy in Figure 1 will be described below. Where in Figure 1 (a) contains 11 nodes to be clustering network; (b) is a candidate chromosomes of this network; (c) is the structure of Figure. Chromosome corresponding candidate solution corresponds network optimal partitioning.

LAR encoding has the following advantages in the detection of network community, the number of communities in complex networks k without specified by automatically determines the number of connected components of chromosome decoding. Therefore, we can the GALS one run in pairs chromosome has a number of different communities to assess the comparison. In addition, the encoding method is very suitable for the standard cross-operator, such as uniform crossover, single-point crossover, multi-point crossover. It is achieved very well for communities merge and split by these cross-operating, and be able to maintain good building blocks to the contrary, if you use traditional string encoding standard crossover operation which will be very easy to destroy the good bricks block, resulting in the failure of the network community detection process.

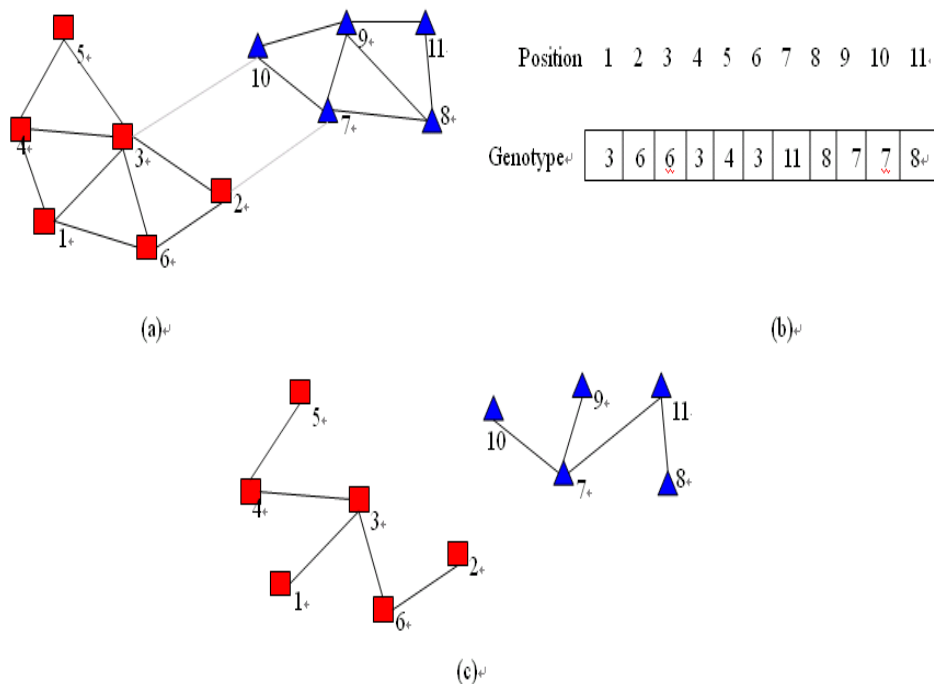


Figure 1. The Adjacent Chromosomal Loci Coding Strategy

2.3. Generating the Initial Population

If the random connection of the method is used to produce the initial individual communicating branch, it corresponds to the figure in the original network may not

communicating, and that it may not be a sub graph of the original network. For example, the i chromosomal loci of an individual corresponding to the value of j , but the edge $\langle i, j \rangle$ in the original network is possible does not exist. However, due to the intuitive prevalent in the community structure of complex network, any node will network with some of its neighbor nodes located in the same community, or their own to form a community, while the other case is unreasonable. Therefore, the solution space of the chromosome can be clipped, for example, the requirements of each individual in the population should correspond to a support of the original network spanning sub graph, and thus will meet the above conditions the individual called security instance. It will be composed of all by a security instance populations known as the safety of the population, from the safety of individuals constitute the solution space is known as the security solution space. LAR encoding the entire solution space is n^n , but security solution space is $\prod_{i=1}^n k_i$. Where n is nodes of the network N , k_i indicates the degree of node i , Since most complex network is sparse graph, k_i can be regarded as a constant ($k_i \ll n$), Therefore, the security solution space is far less than all of the solution space. Clearly, the genetic algorithm limit your search security solution space, you can significantly improve search efficiency and accelerate convergence.

Safety population in order to generate a certain precision, this paper presents a method based on Markov Random Walk. Assuming that an agent along the network N edge set E random walk, it should choose the next step to reach the location of the transition probability between every step. If the agent is currently located in the node i , its next step to reach the transition probability of the neighbor node j is p_{ij} , If the adjacency matrix of the network N is $A = (a_{ij})_{n \times n}$, then

$$p_{ij} = \frac{a_{ij}}{\sum_k a_{ik}} \quad (2)$$

If the matrix representation is used, let x , assuming $D = \text{diag}(d_1, \dots, d_n)$, where $d_i = \sum_j a_{ij}$ said node i of degrees, the probability of the transition matrix $P = (p_{ij})_{n \times n}$ is

$$P = D^{-1}A \quad (3)$$

Higher density within the community connection and community connection between low density complex network community structure are from the point of view of Markov random walk, then a random walk of the agent will be very easy to access its own community where the node, but it is difficult to reach other parts of the network through the boundaries between communities. That is, if the agent from any one of the nodes in the network, the probability that it hovering within their own communities is greater than the probability of out of the community. This article so that any one chromosome for (node) i step transition probability p_{ij} to select the gene values $j \in \{1, \dots, n\}$, Clearly this method can produce a safe initial individual while the individual not only has a certain community divided precision, but also has a strong diversity.

2.4. Selection and Crossover Operator

The selection operator is a global search operator in genetic algorithm. For retaining the best individual in each generation, and accelerating the convergence of the algorithm, this article uses a combination of optimization evolutionary algorithm preference the $\mu + \lambda$ selection strategy [20]. $\mu + \lambda$ strategy is from the parent population (scale μ) and by cross sub-populations variation (scale λ) is the common choice of the largest fitness μ individuals, and as a next-generation parent population. The crossover operator genetic algorithm, a global search operator, so we chose the uniform crossover strategy [21], A

given the two parent individual A and B, and randomly generates a length of n binary vector v , The uniform cross-select which is from the parent individual A is to select 1 gene in vector v , to select 0 gene in v from the parent individual B, thus combined into a new child instance C. if using a mathematical formula, then $C = A.*v + B.*(1-v)$, where point multiplication ($.*$) shows the vectors are multiplied by the corresponding bit. Because individuals are safe in text population initialization method, which means that the value of any gene i is the value of j . It is expressed the edge of $\langle i, j \rangle$ in the original network N from its parent, but also because any offspring genes instance, therefore, the offspring produced by a uniform crossover operation is safe. In addition, because of the uniform crossover and single-point, multi-point cross is different from its chromosome genes. So it is not sensitive sort. The child individual gets any combination of the gene from the parents. Figure 2 for uniform crossover operation has been described. Where Figure 2(a) is the graph of parent individual A and its communicating branch; (b) is the graph of the parent individual B and its communicating branch; (c) parent individuals A and B are using a random binary vector v to get offspring C by uniform crossover; (d) is sub-individual C corresponding graph and its communicating branch.

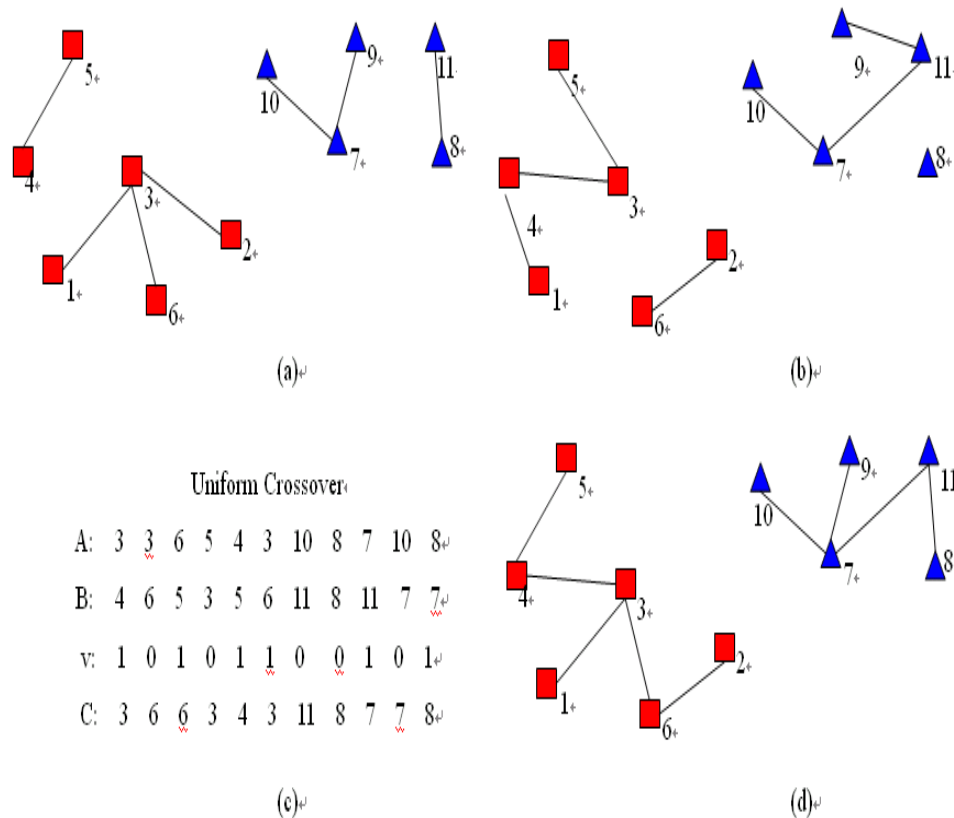


Figure 2. For the Network in Figure 1(a) Shows an Instance of a Uniform Cross-Operation

2.5. Mutation Operator

The mutation operator is the focus of this research-oriented network community detection problem for the lack of traditional mutation operation. This paper first gives a edge chromosomal loci concept, and thus a fast and effective local search variant algorithm oriented for edge chromosomal loci.

2.5.1. Edge Chromosomal Loci: Most of the current LRA coding genetic algorithm with a random mutation method for local search oriented network community detection problem, however, this method is not very satisfactory. Guimera and Amaral think [9], for network community detection problem, the mobile node split, merge and community to community in the current candidate solution is an effective way to generate new candidate solutions. Crossover operator in the genetic algorithm generally regarded as individual macro operation, the mutation operator is regarded as micro-operation of individuals. Genetic algorithm for solving the problem of network community detection crossover operation by splitting and merging of the community is to achieve its global search function mutation operation node through inter-community mobile local search function. This produces a genetic algorithm in order to have a stronger ability to search. This article each LAR chromosome corresponds to a graph, each corresponding to the connected components of a community, it is clear that UC can easily split and merge community crossover operation, mutation operation involved the need for effective realization of inter-community Results point moves, However, our study found that the traditional variation method often cause split or consolidation of the community, cause it can not effectively perform local search function, thereby resulting in the failure of genetic algorithms. For example, suppose that the mutated chromosomal loci i is to select a new assignment j in chromosome g . If node i and node j are previously belonging to different communities, then the variation of behavior is likely caused by the merger of the two communities. The two connected components of that node i and node j may be through the side of the new generation $\langle i, j \rangle$ consolidated into a larger communicating branch. In Figure 3 (c) shown below.

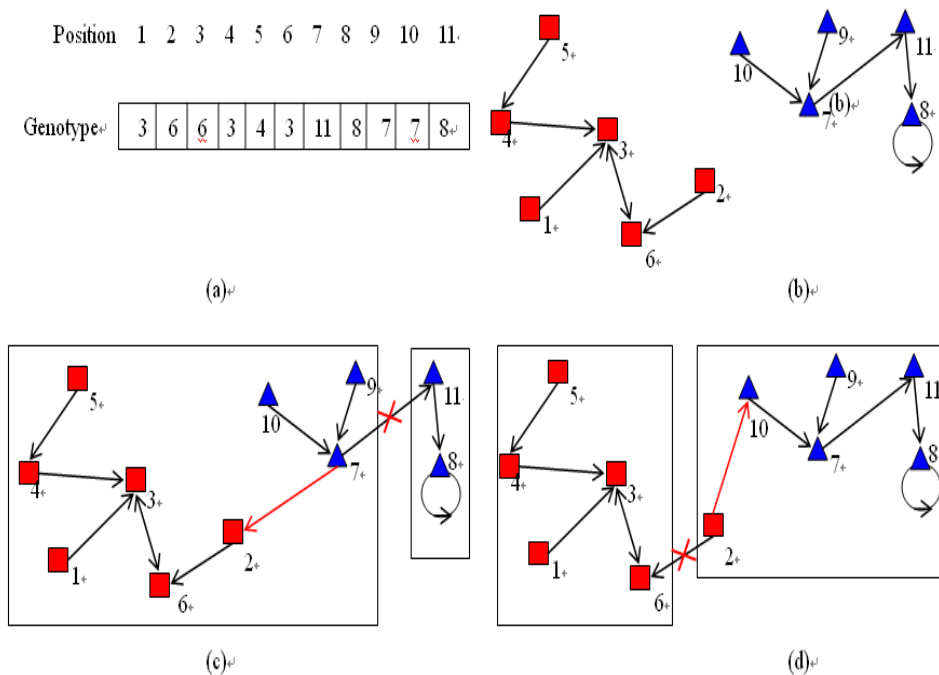


Figure 3. For the Network in Figure 1(a) Shows the Cross-operation Demo

In order to more clearly to describe the process of variation, therefore, it is using digraph instead of no digraph to express the chromosome. Figure 3 (a) is for this network a candidate chromosomes A; (b) is for A corresponds digraph G; (c) a mutation operation is performed in the non-edge node 7, the operation result in the chromosome (A) community split and merge; (d) in the edge node 2 mutation operation, the operation to

achieve the mobile node in the community, and at the same time did not lead to a split or merger of the communities. Obviously this situation is that we do not want to see, which we undertake the following studies.

Definition 2.1 Edge chromosomal loci Take LAR coding any chromosome g , if assignment of all of the chromosomal loci in the g is not for the j , then chromosomal loci j is known as the edge chromosomal loci of chromosome g (edge node).

LAR coding of any chromosome g corresponds to a directed graph G . Definition 2.1 that figure G does not have any junction point to the edge node, so the variation of the edge nodes can achieve single node in the movement between the community, but does not cause merging or splitting of the communities. A simple example is shown in Figure 3 (d). Obviously, the edge gene can be a very good local search function of the genetic algorithm. Now to study edge chromosomal loci is located in the proportion of the chromosome. Randomly generated chromosome g , for example, in this first study, and then extended to general chromosome. Assuming g contains chromosomal loci of n , then the probability of the value j of any one gene i is $p = 1/n$, gene i which is not to be taken the value of j should be $1 - p$. And then the values of all genes are not the value j in the g , so the probability is $\beta = (1 - 1/n)^n$. Therefore the probability of edge chromosomal loci of chromosome g should be β , by mathematical analysis shows that $\beta(n)$ is a monotonically increasing function, and $\lim_{n \rightarrow +\infty} \beta(n) = 1/e \approx 0.3679$, where e is for the natural logarithms. Due to the number of nodes of the complex network is generally greater than 10, so $p(10) = 0.3487$, therefore, the probability of the edge chromosomal loci in chromosome g should be $\beta(n) \in (0.3487 \square 0.3679)$, $n > 10$. Finally, through experimental analysis and found that it is roughly the same proportion of the edge chromosomal loci in normal chromosomes with random chromosome. It would appear, if all of the edge nodes for the chromosome is mutating, then it is equivalent to the mutation rate β for the chromosome mutation, but fortunately, the mutation rate β is not too big or too small, otherwise the advantage of edge chromosomal loci may be weakened.

2.5.2. Mutation Algorithm based on Local Optimization: Based on the concept of edge nodes is proposed in the previous section. Here trying to make a fast and effective local search mutation algorithm for edge node through the analysis of the module function. Effective local search mutation algorithm is given for the movement of the junction point of the edge of the inter-community, and as we are using the module function Q as an objective function, so here first starting from the partial views of each node for the module function Q analysis. The article will explain that the formula (2-1) is converted to the formula (2-4). The function Q is expressed as the sum of the function of all nodes in the network. Obviously the function f can be understood as a fragmentary view of a node from the network to any community view, the actual connection to the difference between desired the number of connections in the case of the number of randomly connected community. The f function of each node in the network can come from a local perspective to evaluate the pros and cons of community structure. Related to the nature of the f function is given below and theorems.

$$Q = \frac{1}{2m} \sum_i f_i, \quad f_i = \sum_{j \in c_r(i)} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \quad (4)$$

Nature 2.1 For $\forall i \in V$, a partial function f_i of the node i in the complex network is associated with it in the community $C_{r(i)}$.

Due to (2-4), the nature of 2.1 is clearly established.

Theorem 2.1 For $\forall i \in V$, if the label of node i will be change under the same premise of the label of other nodes, then the Q function value of the complex network is monotonically increasing with f_i .

On the basis of the above discussion, intuitive phenomenon prevalent in the community structure of complex network, the network of any one node will be located in the same community and some of his neighbor nodes, or their own to form a community. So, in all edge nodes are used as follows mutation strategy, assuming that the set of the neighbors of any edge node i is NS_i . Now we do not need to consider the entire collection V of nodes, only from NS_i to choose a node j as the gene value of the node i is to make local function f_i maximized. Obviously any security individual variation the method is still safe and the individual. Based on this design, a fast and efficient local search mutation algorithm is called LSMA (Local Search based Mutation Algorithm). It is given LSMA description.

Procedure LSMA

Global: g

Begin

Step 1 $C \leftarrow$ for g is decoding

Step 2 For $i=1: n$

Step 3 If i is for edge chromosomal loci

Step 4 $NS_i \leftarrow$ Take all the neighbors of the node i

Step 5 $labels \leftarrow$ Take all nodes corresponding NS_i community label

Step 6 $max \leftarrow -\infty$

Step 7 For each $r \in labels$

Step 8 $f_i \leftarrow$ the value of f function is computed for the label value r of the node i

Step 9 If $f_i > max$

Step 10 $max \leftarrow f_i$

Step 11 $label_i \leftarrow r$

Step 12 End

Step 13 End

Step 14 $g(i) \leftarrow$ take any node of the label $label_i$ in NS_i

Step 15 $C(i) \leftarrow label_i$

Step 16 End

Step 17 End

End

It can be seen. The algorithm LSMA, the change of the tags for each node represents the chromosome g of the junction point between the different communities (or communicating branch) first move. At the same time it can not cause a split or merger of the relevant community. Therefore, only the edge nodes can meet this requirement. In contrast, non-edge node label changes will lead to split or merge existing communities, so they do not meet algorithm LSMA the requirements. So said LSMA mutation algorithm is completely against the edge node design.

It is worth noting that the above variation the algorithm LSMA is using labels synchronous update mechanism. Although the mechanism so LSMA meet any LAR coded chromosome g does not reduce the fitness value LSMA variation are, but also the

risk of the genetic algorithm GALS local optimal solution. If LSMA is using tags asynchronous update mechanism (not perform step 15), then it is equivalent to the genetic algorithm GALS algorithm introduces an adaptive mutation mechanism. GALS algorithm early, since each node opportunity to update their own label, so that the community structure changes in the larger, stronger randomness of the algorithm can be beneficial to its escape from the local optimal solution. GALS algorithm late, since each node updates its own label probability is very small, changes in community structure slowed down, the algorithm almost turned into a greedy search algorithm can be beneficial to its find more accurate global optimal solution. It can be seen that this mechanism and simulated annealing algorithm [9] is somewhat similar to annealing mechanism, it can improve the performance of the genetic algorithm GALS. So this variation the algorithm LSMA final is using tags asynchronous update mechanism.

2.6. Genetic Algorithm GALS Framework

On the basis of the above discussion, this section gives a description of the algorithm GALS:

Procedure GALS

Input: N, L, μ, λ // N represents complex networks, L represents the GALS number of iterations, μ represents population size of the parent, λ represents the offspring population size

Output: C //network community structure

Begin

Step 1 $P \leftarrow$ execute the number of μ , MRW produce initial population

Step 2 For $i=1: L$

Step 3 $P^{(new)} \leftarrow \emptyset$

Step 4 For $j=1: \lambda$

Step 5 $g \leftarrow$ UC is acting on cross the two parent individuals from taking any of P individual

Step 6 $g \leftarrow$ execute LSMA is mutating for g

Step 7 $P^{(new)} \leftarrow P^{(new)} \cup g$

Step 8 End

Step 9 $P^{(u)} \leftarrow P \cup P^{(new)}$

Step 10 $P \leftarrow$ from $P^{(u)}$ to select the best individual of μ // $\mu + \lambda$ selection

Step 11 End

Step 12 $I \leftarrow$ Take largest individual fitness from P

Step 13 $C \leftarrow$ For I is decoding

End

As can be seen, the algorithm GALS is firstly using based on Markov Random Walk (MRW) which is used to generate the initial population. Perform uniform crossover (UC), and then through an iterative local search based mutation algorithm (LSMA) and $\mu + \lambda$ to select the three genetic operators to detect network community structures. Where, UC crossover operator will be through LAR chromosome communities split and merged to achieve its global search function. LSMA mutation operator will make LAR chromosome in the community between the edge nodes for destination mobile to achieve local search function. $\mu + \lambda$ selection operator had although not directly act on a single chromosome, but it is through the "survival of the fittest" selection mechanism to make the high fitness chromosomes into the next generation population, in order to achieve its global search function. In addition, since the initial population produced by the algorithm

MRW is safety populations. While the UC cross and LSMA of change chromosome structure Variability two operators will not undermine the security of the original population. So, algorithm GALS problem in the security solution space search.

3. The Experimental Results

3.1. Experiment

In order to test the performance of the algorithm GALS, the article is using real network test. In the experiment, the algorithm GALS is compared with five kinds of the representative community detection algorithm, where FN, FUA and TGA are based optimization method, where FEC and LPA is based heuristic method. It is worth noting, FUA is regarded as one of the most effective community detection algorithm in its review. TGA is considered to be the most classic of community detection method based on genetic algorithm.

There are three parameters in the algorithm GALS, respectively, as the number of iterations L , parent population size μ and offspring population size λ . They are standard parameters in the genetic algorithm, Based on the literature [14][15][16][21] and some experimental results were set, where $L = 500$, $\mu = 80$, $\lambda = 60$. The algorithm experimental environment processor is Inter(R) core(TM)2 Duo CPU P8700 @2.50GHZ, memory 2.00GB, hard disk 250G, operating system Microsoft Windows XP.

3.2. The Complex Network of the Real World

Real-world network with different topology characteristics usually associated with computer-generated network. Therefore data of seven is widely used in the real network to test the performance of the text algorithm GALS. These data not only contain dozens of small-scale network node, but also has a large-scale network of tens of thousands of nodes. Table 1 shows a simple description of these networks.

Table 1. Real Network Data

Networks	V(N)	E(N)	Description
karate	34	78	Zachary's karate club [4]
dolphin	62	160	Dolphin social network [5]
polbooks	105	441	Books about US politics [10]
Football	115	613	American College football [3]
jazz	198	5484	Jazz musicians network [22]
email	1133	5451	Email network of human interactions [23]
internet	22963	48436	A snapshot of the Internet by Newman [24]

As a real-world network community structure is usually unknown, so here recognized module function Q as a different algorithm clustering quality metrics. Real network in Table 1, Table 2 shows that these algorithms are running 50 times the average Q - function value. It can be seen that for these real network, the quality of clustering algorithm GALS is less than the FUA, but is superior to the other four algorithms.

Table 2. Algorithm GALS on a Real Network and Other Algorithms Clustering Quality Comparison

Q-value	FN	FEC	LPA	FUA	TGA	GALS
karate	0.3807	0.3744	0.3646	0.4188	0.4039	0.4198
dolphin	0.5104	0.4976	0.4802	0.5268	0.5241	0.5294
polbooks	0.5020	0.4904	0.5006	0.4986	0.5245	0.5272
Football	0.5497	0.5697	0.5865	0.6046	0.5937	0.6046
Jazz	0.4389	0.4440	0.3422	0.4431	0.4406	0.4449
email	0.5037	0.5173	0.3706	0.5406	0.1871	0.5599
internet	0.6378	0.6104	0.4978	0.6613	0.1141	0.6560

For example, the karate club network (karate) is Zachary built by two years of observations of an American University karate club [4]. It is 34 members in the club as a node. If the friendship relationship exists between the two members, between their corresponding vertex will have one side connected. Later, due to differences of opinion, the club eventually is split into two new club managers and coaches as the core. True community structure of the network is shown in Figure 4. Nodes of the same shape of the same color belong to the same real communities. The compact node heap is algorithm GALS a community. Node 1 on behalf of the club coach, Node 33 managers on behalf of the club, after splitting the left all the red square node represents the club members in the same group with the coach. After the division of the right side all blue triangle node behalf of the club members and managers in the same group. The network, for example random running time algorithm GALS, the results obtained contains four communities. As shown in Figure 4, you can see that the algorithm GALS not only completely correct detection karate true community structure of the network, but also two real community subdivided became the two internal closely sub-communities. In addition, with network running 50 times for karate, algorithm GALS get the average Q function value of 0.4198, which is bigger than real community structure of the network corresponding to the Q value 0.3715.

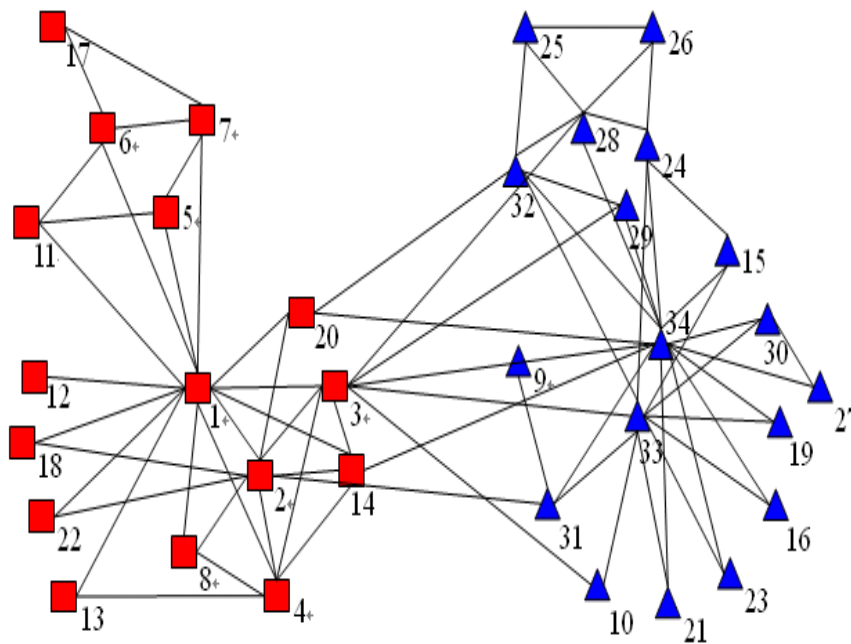


Figure 4. The GALS on the Karate Network Clustering

4. Conclusion

Proposed in this paper based on local search genetic algorithm GALS is modularity function Q as the objective function, is encoding by LAR. The algorithm is using Markov random walk method to generate the initial population, and then executed by iteration uniform crossover, local search variation and $\mu+\lambda$ choose three genetic operators to detect network community structures. By testing on a real network, we can found experimental results verify the effectiveness and efficiency of the algorithm and had compared with the current representative of the community-detection algorithms.

References

- [1] D. J. Watts and S. H. Strogatz, "Nature", vol. 393, (1998), pp. 440-442.
- [2] A.-L. Barabási and R. Albert, "Science", vol. 286, (1999), pp. 509-512.
- [3] M. Girvan and M. E. J. Newman, "Proceedings of National Academy of Science", vol. 9, (2002), pp. 7821-7826.
- [4] W. W. Zachary, "Journal of Anthropological Research", vol. 33, (1977), pp. 452-473.
- [5] D. Lusseau, "Proceedings of the Royal Society B: Biological Sciences", vol. 270, (2003), pp. 186-188.
- [6] B. Yang, D. L. Liu and H. Ma, "Journal of Software", vol. 20, (2009), pp. 54-66.
- [7] M. E. J. Newman and M. Girvan, "Physical Review E", vol. 69, (2004), p.113.
- [8] M. E. J. Newman, "Physical Review E", vol. 69, (2004), p. 133.
- [9] R. Guimera and L. A. N. Amaral, "Nature", vol. 433, (2005), pp. 895-900.
- [10] M. E. J. Newman, "Modularity and community structure in networks", Proceedings of National Academy of Science, vol. 103, no. 23, (2006), pp. 8577-8582.
- [11] V. D. Blondel, J. L. Guillaume, R. Lambiotte and E. Lefebvre, "Journal of Statistical Mechanics: Theory and Experiment", 10008, (2008).
- [12] M. J. Barber and J. W. Clark, "Physical Review E", vol. 80, 026129, (2009).
- [13] M. Tasgin, A. Herdagdelen and H. Bingol, "Community detection in complex networks using genetic algorithms, (2007).
- [14] C. Pizzuti, "Community detection in social networks with genetic algorithms", Proceedings of Genetic and Evolutionary Computation Conference (GECCO'08), (2008) Atlanta, Georgia, USA: ACM Press.
- [15] C. Pizzuti, "A multi-objective genetic algorithm for community detection in networks", Proceedings of the 21th annual IEEE International Conference on Tools with Artificial Intelligence (ICTAI'09), (2009) Washington, DC, USA.
- [16] C. Shi, Z. Yan, Y. Wang, Y. Cai and B. Wu, "Advances in Complex Systems", vol. 13, (2010), pp. 3-17.
- [17] Y. Park and M. Song, "A genetic algorithm for clustering problems", Proceedings of the 3rd Annual Conference on Genetic Programming, (1998) Madison, Wisconsin, USA.
- [18] J. Handle and J. Knowles, "IEEE Transactions on Evolutionary Computation", vol. 11, (2007), pp. 56-76.
- [19] T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, "Introduction to Algorithms", Cambridge, MA: MIT Press (2001).
- [20] E. M. Montes and C. A. C. Coello, "IEEE Transactions on Evolutionary Computation", vol. 9, (2005), pp. 1-17.
- [21] G. Syswerda, "Uniform crossover in genetic algorithms", Proceedings of the 3rd International Conference on Genetic Algorithms, Fairfax, Virginia, (1989) USA: Morgan Kaufmann.
- [22] P. M. Gleiser and L. Danon, "Advances in Complex Systems", vol. 6, (2003), pp. 565-573.
- [23] R. Guimerà, L. Danon, A. Diaz-Guilera and F. Giralt, "Physical Review E", vol. 68, 065103, (2003)/
- [24] Network data from Mark Newman's home page, 2006, <http://www.personal.umich.edu/~mejn/netdata>.