2020 19th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)

A Discrete Random Drift Particle Swarm Optimization with Modularity in Community Detection

1st Fen Wang
School of Internet of Things Engineering
Jiangnan University
Wuxi, China
1401711548@qq.com

2nd Li Sun
School of Internet of Things Engineering
Jiangnan University
Wuxi, China
lisun@jiangnanedu.cn

3rd Jun Sun
School of Internet of Things Engineering
Jiangnan University
Wuxi, China
sunjun_wx@hotmail.com

3rd Qidong Chen
School of Internet of Things Engineering
Jiangnan University
Wuxi, China
cqd_jnu@hotmail.com

***Abstract*—In the field of complex networks, community detection is one of important research objects. To solve the problem of poor quality and the unstable result with community structure, we propose a community detection optimization algorithm based on random drift particle swarm optimization (RDPSO) algorithm (DRDPSO-net), in which we use discrete method to update the network information. Through the discrete particle evolution process and local greedy strategy with network topology character, DRDPSO-net can obtain a better quality of community division. In addition, several representative real networks are used to verify the performance of DRDPSO-net. By comparing them across several algorithms, DRDPSO-net has more desirable value among those algorithms. Furthermore, the experimental results demonstrated that DRDPSO-net obtain a valid and steady community structure.**

Keywords—community detection, DRDPSO-net, modularity, network topology information

I. Introduction

In the complex network, community detection can reveal the topological characteristics in the network, which are more sparse connections among communities and more close connections between communities. Many important tasks are based on it, including network virus transmission, link prediction, network evolution analysis, and graph mining, etc. Therefore, the algorithm of community detection has been the hot research topic in various fields.

At present, according to different research perspectives, we can solve this problem in community division by clustering or optimzation ways. However, the classical methods are GN[1], Infomap[2]etc., which are classified as the clustering algorithm in community detection. Based on the optimization perspective, Louvain is a representative algorithm, but most algorithms are based on a heuristic evolutionary algorithm to gradually optimize the modularity function and obtain the community structure with the maximum modularity in recent years. Since 2008, Pizzuti has successfully applied the GA algorithm to obtain community structure[3]. Subsequently, many evolutionary algorithms have also been used in community division, such as IMAs[4], and so on. In order to solve the limitation of the resolution of single target modularity, those MOGA-@net[5]algorithms are collectively known as the multiobjective optimization algorithm in community detection. Through the combination of different community functions, its Pareto optimal method can alleviate the defects of single-objective modularity optimization. But it increases the complexity of the algorithm and running costs. On the other hand, there are also some challenging problems in evolutionary algorithms of community detection, such as too many parameters and low convergence. And the complexity of cross variation leads to the instability of community structure. The intelligent particle swarm optimization has few parameters, strong convergence, and simple operation1. Cai introduced a new discrete formula with PSO algorithm for signed network, which was effective and promising to community structure. However, this method war poor performance in the searching ability of the algorithm. Cao proposed the improved simple discrete PSO algorithm (ISPSO) and improved discrete PSO with a reddened operator (IDPSO-RO) Algorithm. And a community correction strategy for isolated nodes is proposed to optimize the results. But the partitioning result is not very effective for the community modularity. Chaitanya[6] proposed a method utilizing particle swarm optimization and dynamic neighborhood topology in community detection. Besides, optimization algorithms based on other animal behavior are introduced into community discovery, such as intelligent bee colony algorithm , bat optimization algorithm , etc.

Although community detection based on the framework of optimization algorithms has achieved different levels of success, it still has some problems that the second-best scheme is replacing its best project. So, in order to overcome these problems and improve the quality of community structure, we proposed a new algorithm for community detection. In PSO algorithms, the random drift particle swarm(RDPSO)[7] algorithm has strong global search ability. And it can avoid local optimization. Our work in this paper is that redefine the particle evolution strategy combining the network topology information for complex network, and the optimal community partition quality can be obtained.

The rest of this paper is organized as follows. Section II focuses on the related concept of community detection and instructions of the RDPSO algorithm; Section III shows the initialization and discretization of RDPSO and the processes of DRDPSO-net; Section IV is the experiment results in a

2473-3636/20/$31.00 ©2020 IEEE

DOI 10.1109/DCABES50732.2020.00070

synthetic network and real networks; Finally is the conclusion and prospect.

II. Related Background

A. The Community Detection Problem

In graph theory, the relationship between groups can be defined by a mathematical formula. A network can be represented by two-tuple $G = (N, E)$, where $N=\{n_1,...,n_n\}$ is nodes in network, and $E=\{(I, J)|n_i\in N, n_j\in N, \text{ and } i\neq j\}$ is edges in network. The task of community detection is to obtain the same nodes of graph to a group, which are also called community. Let $C=\{C_1, C_2,...C_n\}$ be communities in network, then the ith's community must be satisfies the following conditions:

$$C_i\subset V \text{ and } C_i\neq\emptyset,\ i=1,2...,n$$

$$C_i\neq C_j, \forall\ i\neq j \text{ and } i,j\in\{1,2...,n\}$$

$$\bigcup_{i=1}^{n} C_i=V$$

B. Community Detection Optimization Problem

For the community detection optimization problem, modularity function is used to evaluate the community division standard. The article introduces the physical explanation of modularity[8]. Therefore, the target function of the optimization community detection problem can be defined as:

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

Here, m is the number of edges in the network; $A_{i,j}$ is an element of the adjacency matrix; k_i and k_j represents the degree of network node i and j; $\delta(i,j)=1$ means that node i and node j are in the same community; otherwise the value equals to 0. A larger Q value represents the community structure is more obvious.

III. Proposed DRDPSO-net Algorithm

In community detection of the complex network, the modularity optimization has become a significant object to solve the problem. Therefore, the straightforward idea is to use RDPSO algorithm to find the maximum modularity. However, RDPSO algorithm performs better on continuous functions rather than on discrete optimization problems. Since the network data is generally regarded as discrete data, hence we propose a DRDPSO-net algorithm for complex network clustering.

DRDPSO-net algorithm divide into two main steps, including initial particle operation in section A and the particle updating operation in section B and C.

A. Particle Coding and Preprocessing

In DRDPSO-net algorithm, particle coding changes the position of the particle. Here we use a straightforward representing scheme, as shown in Fig 1. The position permutation value in Fig.1 is an n-dimensional vector $X_i=\{x_{i,1},x_{i,2},...x_{i,j},.,x_{i,n}\}$, where n is number of nodes of the whole network, each $x_{i,j}$ value can be a random integer between 1 and n, which represents the community number of jth node. So, we can understand that the position vector represents a solution to the community division.

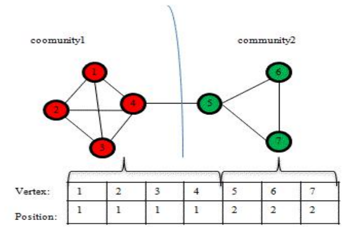

Fig. 1. Particle coding representation graph

Here, we apply the neighbor node's information to divide a preliminary community structure. This initial method, based on the membership of neighbor nodes, get those nodes with more close and more tight connection. It can be easily divides same community, and can find the potential community structure in the network, which helps to improve the accuracy of the experimental result.

B. Proposed the DRDPSO-net algorithm

Here, we explain the DRDPSO-net algorithm by redefining the particle status formula in the traditional algorithm. In the DRDPSO-net algorithm, the position vector of the particle is mapped to the community number that the nodes belong to in the network. And the velocity of particles is regarded as an updated judgment variable.

The equation in DRDPSO-net algorithm is defined in detail as follows:

$$v_{i,j}^{t+1}=\alpha\otimes\left|C_{i,j}^{t}\ominus x_{i,j}^{t}\right|\otimes\varphi_{i,j}^{t}\oplus\beta\otimes\left(p_{i,j}^{t}\ominus x_{i,j}^{t}\right) \tag{2}$$

$$x_{i,j}^{t+1}=x_{i,j}^{t}\odot v_{i,j}^{t+1} \tag{3}$$

where the α is the thermal coefficient; β is the drift coefficient and φ is random number generated by $N\sim(0,1)$; the $p_{i,j}^{t}$, $C_{i,j}^{t}$ are the part of the drift motion and thermal motion of particle in the canonical RDPSO algorithm [14]. Here are redefined by:

$$p_{i,j}^{t}=\begin{cases}P_{i,j}^{t} & fi<0.2\\ G_{i,j}^{t} & 0.2<fi<0.4\\ P_{i,j}^{t}\cup G_{i,j}^{t} & fi>0.4\end{cases} \tag{4}$$

$$C_{i,j}^{t}=\left(\sum_{1}^{n}P_{i,j}^{t}\right)\%n \tag{5}$$

and the fi is a random number generated by $U\sim(0,1)$.

In addition, the equation 2,3 are the discrete velocity and position. We use '$\oplus$', '$\ominus$',' '$\otimes$' and '$\odot$' symbols to represent the discrete operations. It makes the velocity as binary variable. Hence, we describe the processing about discrete velocity in the equation 2 in detail:

$$v_{i,j}^{t+1}=\begin{cases}0 & random>Vd_{i,i}^{t+1}+Vr_{i,j}^{t+1}\\ 1 & random<Vd_{i,j}^{t+1}+Vr_{i,j}^{t+1}\end{cases} \tag{6}$$

where the $\mathrm{Vd}_{i,i}^{t+1}$, $\mathrm{Vr}_{i,j}^{t+1}$ are the drift velocity and thermal velocity. The sum of them is representing the probability of updating the velocity. They are also changed by:

$$vd_{i,j}^{t}=\begin{cases}\beta*vd_{i,j}^{t} & x_{i,j}^{t}\notin p_{i,j}^{t}\\ 0 & x_{i,j}^{t}\in p_{i,j}^{t}\end{cases} \tag{7}$$

$$vr_{i,j}^{t}=\begin{cases}\alpha * vr_{i,j}^{t} * \varphi_{i,j}^{t} & x_{i,j}^{t} \neq C_{i,j}^{t}\\ 0 & x_{i,j}^{t}=C_{i,j}^{t}\end{cases} \quad (8)$$

In the equation 3 is the discrete operation of the swarm position. Base on the result of velocity update, the position of particle is updated by:

$$x_{i,j}^{t+1}=\begin{cases}x_{i,j}^{t} & v_{i,j}^{t+1}=0\\ maxfit & v_{i,j}^{t+1}=1\end{cases} \quad (9)$$

If $v_{i,j}^{t+1}=0$, the jth component of particle i remain the same; otherwise, the $x_{i,j}^{t+1}$ is set to maxfit that is updated according to the community number by the greedy local search strategy. The process of updating position is described in the next section.

C. Greedy local search strategy

For the position updating of the particle is a key step in the algorithm. This greedy selection in the strategy is based on the local modularity incremental function. And the function can be used to detect the topology performance between two communities.

In the processing of particle position updating, the local search firstly finds community numbers that neighbor nodes of the particle *i* belong to. Then let the particle *i* join in one of the neighbor nodes community and calculate them into function. Finally, the position of particle *i* is set to the neighborhood community number that maximizes local modularity function.

D. Processing of the Proposed Algorithm

Through a combination of the preprocessing strategy and greedy local research strategy in DRDPSO-net algorithm, the obtained community structure is more valid. The procedure of the DRDPSO-net algorithm is outlined below:

Algorithm DRDPSO-net

Input: Adjacency matrix of complex networks *G*, population size *pop* and maximum generation *maxgen*

Output: communities

Step 1 initialization:

1. The position vector X^0 according to the section III-A;
2. The velocity vector V^0 select a number in $\{0,1\}$ at random;
3. Calculate the fitness of each particle position X^0 ($X\_fitness^0$) according to Eq.(1), the personal best solution $P=X^0$, $P\_fitness=X\_fitness^0$, select the current global best solution *G*, *G_fitness* from *P*, $\boldsymbol{P\_fitness}$;

Step 2 Updating:

4. For *t in range*(*1,maxgen*):
5. Update the current velocity V^t,position $X^t$and fitness $X\_fitness^t$:according to the Eq.(2-9) and local search in section III-C;
6. Update the *P* particle and *G* particle, according to this process: if $X\_fitness^t > P\_fitness$, then $P=X^t$ and $P\_fitness=X\_fitness^t$, G set the current particle by rank all swarm from P;
7. When the stop condition is satisfied, it output the corresponding partition of the group best particle,

IV. Experiments and Results

In this section, we obtain experimental results by conducting experiments with DRDPSO-net algorithm by comparing it with five algorithms on the synthetic networks. The 5 algorithms are GN(A1), Louvain(A2), Infomap(A3), IDPSO-RO(A4), BSOCD(A5) algorithm. All the experiment results are obtained by conducting 30 times independently.

The experiments are implemented on a system with Intel(R) Core(TM) i7 CPU @2.5GHz and 8G of memory. And the code runs in the python 3.6 and use the networkx API.

A. Comparison Results on Real network

Here we applied four real networks in experiments. Those networks consist of different numbers with nodes and edges . Their information is shown in table I.

TABLE I. Real Network

Dataset	*Nodes*	*edges*
Karate	34	78
Dolphin	62	159
Football	115	613
Polbooks	105	441

We have recorded the maximum modularity(*Qmax)*, and the corresponding average modularity(*Qavg)* in Table II and Table III. The Table II shows results on the karate and dolphin networks. It can see that the DRDPSO-net get the community structure with max modularity. Meanwhile, football and polbooks networks can get the Qmax=0.6044 and Qmax = 0.5269 in Table III. Their community structure are more steady than other algorithms. In all, the DRDPSO-net can performs very well in terms of the best optimal result and stable community structure among six algorithms.

TABLE II. The Value of Modulaity on Karate and Dolphin

	karate		dolphin	
	Qmax	*Qavg*	*Qmax*	*Qavg*
A	0.4013	-	0.5194	-
A2	0.4188	0.3994	0.5188	0.5123
A3	0.402	0.402	0.5277	0.5184
A4	0.402	0.3134	0.5253	0.4719
A5	0.4174	0.3897	0.5246	0.5046
Our algorithm	0.4198	0.4198	0.5285	0.5283

TABLE III. The Value of Modulaity on Football and Polbooks

	football		polbooks	
	Qmax	*Qavg*	*Qmax*	*Qavg*
A	0.5996	-	0.5168	-
A2	0.6019	0.6019	0.5262	0.5040
A3	0.6005	0.6005	0.5228	0.5228
A5	0.5998	0.5869	0.5262	0.5040
A4	0.6006	0.5862	0.5267	0.5170
Our algorithm	0.6044	0.6044	0.5269	0.5269

B. Community Struture of real network on DRDPSO-net

We use a graphical tool Gephi to show the best results obtained by DRDPSO-net. The Fig.2-Fig.5 shows the community division with max modularity. In Fig2, the karate network is divided into four communities. Although this structure is not true network partitioning, but those nodes of blue community in this structure are just separated from green community and those nodes of purple community are separated from red community. And in [6] proposed that the best modularity does not correspond to the true network partitioning. The dolphin, football, polbooks network are divided into 5, 9, 4 communities respectively. In general, the community structure is reliable and indicative in real nertworks.

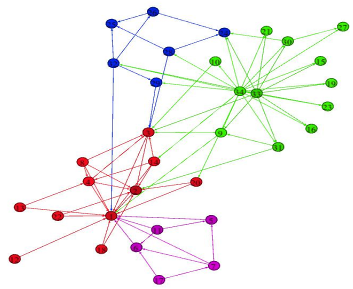

Fig. 2. Karate community structure with Qmax = 0.4198

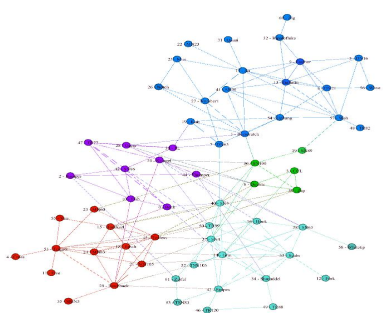

Fig. 3. Dolphin community structure with Qmax = 0.5285

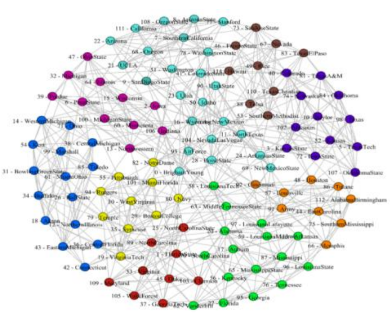

Fig. 4. Football community structure with Qmax = 0.6044

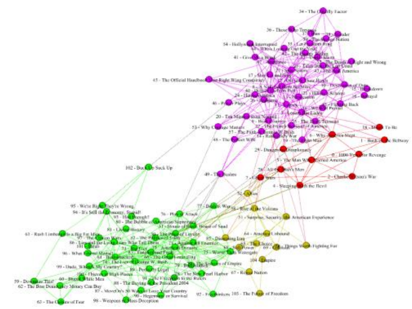

Fig. 5. Polbooks community structure with the Qmax =0.5269

CONCLUSION

In this paper, we proposed the DRDPSO-net algorithm for community detection. It can automatically detect the number of communities without prior knowledge and find the optimal global value. Besides, we use the preprocessing strategy and greedy local search strategy to get a high-quality community. In the experiments, the results show the DRDPSO-net algorithm is validity and efficiency. Although the algorithm has a high value of modularity in the real network.

However, the algorithm can not obtain satisfactory community structure on a large network with the big averaged degrees of the nodes. Therefore, we will pay attention to deal with this problem in the future.

REFERENCES

[1] Girvan M , Newman M E J . Community structure in social and biological networks[J]. PNAS, 2002, 99.

[2] Rosvall M, Bergstrom C T. Maps of random walks on complex networks reveal community structure[J]. Proceedings of the National Academy of Sciences, 2008, 105(4): 1118-1123.

[3] Pizzuti C. Ga-net: A genetic algorithm for community detection in social networks[C]//International conference on parallel problem solving from nature. Springer, Berlin, Heidelberg, 2008: 1081-1090.

[4] Gong M, Cai Q, Li Y, et al. An improved memetic algorithm for community detection in complex networks[C]//2012 IEEE Congress on Evolutionary Computation. IEEE, 2012: 1-8.

[5] Pizzuti C, Socievole A. Multiobjective Optimization and Local Merge for Clustering Attributed Graphs[J]. IEEE transactions on cybernetics, 2019.

[6] Chaitanya K, Somayajulu D, Krishna P R. A PSO Based Community Detection in Social Networks with Node Attributes[C]//2018 IEEE Congress on Evolutionary Computation (CEC). IEEE, 2018: 1-6.

[7] Sun J, Wu X, Palade V, et al. Random drift particle swarm optimization algorithm: convergence analysis and parameter selection[J]. Machine Learning, 2015, 101(1-3): 345-376.

[8] Clauset A, Newman M E J, Moore C. Finding community structure in very large networks[J]. Physical review E, 2004, 70(6): 066111.

[9] Belkhiri Y, Kamel N, Drias H, et al. Bee Swarm Optimization for Community Detection in Complex Network[C]//World Conference on Information Systems and Technologies. Springer, Cham, 2017: 73-85.