# A Personalized Recommendation Method of Social Network Based on Multi Label and Association Rules

Zixuan WANG and Zongfeng ZOU

*Department of Information Management,* School of Management, Shanghai University, Shanghai, 200444, China

***Abstract* — In order to improve the effect of signature network community discovery, and to solve the problem of great limitation on the optimization for the single index of the network community which is caused by data coupling and dependence of the evaluation index, a kind of FS-UD-PSO signature network community discovery algorithm which is based on location repair and particle replacement is proposed. Firstly, the signature network model is studied, and the evaluation index of the signature network community and the multi-objective Pareto optimization objective model are given under the premise of considering the data coupling and dependence; secondly, multi-objective optimization particle coding and updating rule for the signature network model is constructed, and the location repair and particle replacement are designed according to the characteristics of the signature network; at the same time, in order to improve the performance of multi-objective particle swarm optimization algorithm, the multi-objective particle swarm optimization algorithm (FS-UD-PSO) with quick sorting and uniform density is designed; finally, through comparing with the standard test set, the effectiveness of the proposed FS-UD-PSO signature network community is verified.**

***Keywords - community discovery; multi-label; data clustering; particle swarm optimization algorithm; filtering algorithm***

## I. INTRODUCTION

Omnipresent network is changing our daily life in an unprecedented way [1~2]. It has important significance in studying these complex networks from the aspects of theory analysis and practical application. The direct method of analyzing complex networks is that the network is represented by a graph. The graph consists of a set of nodes and edges [3]. The nodes represent the network objectives, and the edges represent the relationships between objectives. The characteristics of network can be obtained by analyzing the characteristic of graph. Network has many characteristics, such as worldlet, scale-free and so on, in which the network community structure is the most popular characteristic. In the graphic language, a community is called a sub-graph, which is characterized by the similarity of the nodes in the community to be higher than the similarity between the communities [4]. Although the discovery efficiency of single-objective community is good and can obtain the specific discovery of a specific network community, in real cases, however, the network community process needs to meet multiple requirements, and such objectives will be conflictive with themselves and the community using single objective is found to be inconsistent with the actual situation, so community discovery using multi-objective evolutionary algorithm is concerned by scholars [11]. In the literature [12], the evolutionary algorithm and programming formula mode are used to obtain the synchronization optimization of internal and external links of the network community, but the computation of the algorithm is too complex. In the literature [13], multi-objective PSO algorithm (MODPSO) in discrete state is designed and community clustering discovery of complex network is achieved, but the algorithm has defects in maintaining the ability of search diversity, which is not conducive to obtain the second-best solution set of overall Parato.

This paper takes the signature network community discovery as the study object and a community discovery method based on multi-objective particle swarm optimization is proposed. The main contributions are as follows: (1) a new problem model with multi-objective optimization for signature network community discovery is constructed based on the link density of nodes; (2) according to the characteristics of the signature network, the multi-objective optimization particle coding and updating rules of the model are constructed, and the corresponding position repair and particle replacement are designed; (3) in order to improve the performance of multi-objective particle swarm optimization algorithm, the multi-objective particle swarm optimization algorithm with quick sorting and uniform density is designed;

## II. PROBLEM DESCRIPTION

### A. Signature Network Model

The network community is usually represented as a directed graph consisting of nodes and edges. Edges can be divided as 2 types of the positive and the negative. The detection task of network community is to divide the symbol network into different clusters according to some principles. Each cluster is usually called as community.

Definition 1: (network community) as for undirected linked network, definite structure $G=(V,E)$, where $V$ is the network vertex (node) and $E$ is the network edge set, and $E=\{e=(u,v),u\in V,v\in V\}$ ; $G$ can be represented by the matrix A with the size of $|V|\times|V|$ . If the network edge

$e=(i,j)\in E$, then $A_{ij}=1$; If the frontier of the network is $e=(i,j)\notin E$, then $A_{ij}=0$; community structure discovery is $P=(V_1,V_2,\cdots,V_m)$, which is the $m$ cluster division of vertex set $V$; vertex $V_i$ satisfies the following limits: $V_i\in V$, $V_i\neq\varnothing(i=1,2,\cdots,m)$, $\bigcup_{i=1}^{m}V_i=V$ and $V_i\cap V_j=\varnothing(i\neq j)$.

Definition 2: (signature community model) given the network model $G=(V,PL,NL)$, where $V$ is the network vertex (node) and $PL$ and $NL$ are respectively the positive and the negative link sets. Supposed that $A$ is the adjacent matrix of $G$ and $l_{i,j}$ is the link between the node $i$ and $j$.

According to definition 1 and definition 2, it is known that the difference between the signature network model and the network community is that the former is divided into positive and negative links, and it is reflected in the model. Fig. 1 shows an example of a signed network with 9 nodes and 16 edges.

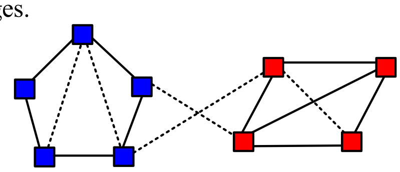


Figure 1. Signature network example

As for sub-graph $S\subset G$, the node $i\in S$. Supposed that $(d_i^+)^{in}=\sum_{j\in S,l_{ij}\in PL}A_{ij}$ and $(d_i^-)^{in}=\sum_{j\in S,l_{ij}\in NL}|A_{ij}|$ are respectively the positive and negative free degrees of node $i$, the condition for $S$ to be the strong signature community is that:

$$\forall i\in S,(d_i^+)^{in}>(d_i^-)^{in} \tag{1}$$

Supposed that $(d_i^+)^{out}=\sum_{j\notin S,l_{ij}\in PL}A_{ij}$ is the positive external free degree and $(d_i^-)^{out}=\sum_{j\notin S,l_{ij}\in NL}|A_{ij}|$ is the negative external free degree, the condition for $S$ to be the weak signature community is that:

$$\begin{cases}\sum_{i\in S}(d_i^+)^{in}>\sum_{i\in S}(d_i^+)^{out}\\ \sum_{i\in S}(d_i^-)^{out}>\sum_{i\in S}(d_i^-)^{in}\end{cases} \tag{2}$$

In the weak signature community, the positive link between the community internal nodes and the negative link between the community nodes are all existed in density. According to Fig. 1, we can see that the task of the signature network community discovery is to divide the whole network into a number of communities, but the division standard is still a difficult problem.

## *B. Index Evaluation of Signature Network Community*

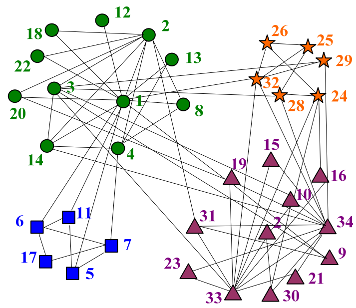
In order to give the quantitative standard for the signature community structure, literature [14] proposes a new module quantization standard; the new index $SQ$ is called as the signature modularization:

$$SQ=\frac{1}{2w^{+}+2w^{-}}\sum_{i,j}(w_{ij}-(\frac{w_i^+w_j^+}{2w^+}-\frac{w_i^-w_j^-}{2w^-}))\delta(i,j) \tag{3}$$

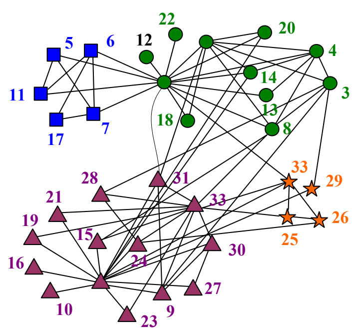
In the formula, $w_{ij}$ is the weight of the signature adjacent matrix and $w_i^+(w_i^-)$ is the total of all the positive (negative) weights. If the node $i$ and $j$ is in the same team, $\delta(i,j)=1$. Otherwise, $\delta(i,j)=0$. The greater the value of $SQ$ is, the better the separation degree of community structure is. At the same time, another silhouette index is that:

$$GS=\sum_{i=1}^{m}\frac{d^{out}(V_i)}{2d^{in}(V_i)+d^{out}(V_i)} \tag{4}$$

Fig. 2 shows the signature community structures under two conditions when the iterations are 90, $SQ=0.423$, $GS=0.698$ and when the iterations are 110, $SQ=0.763$, $GS=0.699$.



(a) Structure 1



(b) Structure 2

Figure 2. Signature community discovery

According to Fig. 2, it can be known that with the optimization of the signature community discovery process, the $SQ$ value of the signature module index is also increased correspondingly, but the value of $GS$ index does not appear to increase.

As for signature network $G=(V,PL,NL)$, $|V|=N$; the function of signature community discovery is that $V\to P$,

DOI 10.5013/ IJSSST.a.17.40.01

ISSN: 1473-804x online, 1473-8031 print

$|P| = 2^N$ ; evaluation vector of community discovery is that $F = (F_1, F_2, \cdots, F_k)$; element $F_i : p \to R$ $(p \in P, i = 1, \cdots, k)$. The nature of evaluation criteria for the signature community is as follows:

Firstly, the definition of coupling is given, and the signature network $G$ is in the process $(m > n)$ of iteration $t_n \to t_m$ for the function of community discovery; if the evaluation vector element $F_i$ of the community discovery meets the condition, the coupling between $F_i$ and $F_j$ is that:

$$F_i(t_m) > F_i(t_n), \quad \exists F_j(t_m) < F_j(t_n) \wedge j \neq i \tag{5}$$

Secondly, the definition of dependence is given; if the priori value of signature network $G$ of community discovery is $P_1^{true}$, using the optimization evaluation criteria $F_i$ and $F_j$ to conduct the division of the signature network $G_1$, it can be obtained that network communities are $P_1^i$ and $P_1^j$ ; $F-Measure(P_1^{true}, P_1^i) > F-Measure(P_1^{true}, P_1^j)$, and it can be expressed as $F_i \overset{G_1}{\succ} F_j$; at the same time, as for the signature network $G_2$ of the priori structure $P_2^{true}$ of the community, it has $F_j \overset{G_2}{\succ} F_i$, so the dependence exists between $F_i$ and $F_j$.

Compound the evaluation criteria based on the weight priori knowledge so as to conduct optimization based on the single-objective algorithm. However, there is coupling between different evaluation indicators, which leads to the non-ideal combined effect, and the optimal solution can not be obtained.

### C. Optimization Objective

The above definition shows that the model optimization needs to adopt multi-objective evolutionary algorithm. Here, the optimal relevant definition of multi-objective Pareto is given.

(1)Pareto domination: as for the process $V \to P$ of signature network model discovery, and the two sorts of community forms in $G = (V, PL, NL)$ is $P_1, P_2 \in P$ ; when the following conditions are met, it can be regarded that $P_1$ dominates $P_2$, namely $P_1 \succ P_2$.

$$\forall i \in (1, 2, \cdots, n): \quad F_i(P_1) \geq F_i(P_2) \tag{6}$$

(2)Pareto equivalence: when the following conditions are met, it can be regarded that $P_1$ is equivalent to $P_2$, namely $P_1 = P_2$.

$$\forall i \in (1, 2, \cdots, n): \quad F_i(P_1) = F_i(P_2) \tag{7}$$

(3)Pareto unknown: when the following conditions are met, it can be regarded that the relation between $P_1$ and $P_2$ is unknown, namely $P_1 \Diamond P_2$.

$$\neg(P_1 \succ P_2) \wedge \neg(P_2 \succ P_1) \tag{8}$$

(4)Pareto optimization: when the following conditions are met, a certain strategy $P^*$ inside the strategy set $P$ of the signature network model of the community discovery is the optimal strategy of Pareto.

$$\neg \exists P_i \in P : P_i \succ P^* \tag{9}$$

Then, the above strategy constructing set $PS = \{P^* | \neg \exists P_i \in P : P_i \succ P^*\}$ is the optimal strategy set for the signature network model of the community discovery.

(5)Pareto frontier: the evaluation vector set of the optimal strategy for the signature network model is the Pareto community frontier

As for signature network $G = (V, PL, NL)$, $F$ is the predefined target set and the community division number is $k$ ; this value can be given according to the evolutionary algorithm for clustering or predefining, then the community discovery process is to search $F$, which makes $P^* = \arg max_{P_i \in P} F(P_i)$ as the optimal division. According to the above definitions, the multi-objective community discovery of signature network model can be defined as:

$$\max F(x) = (F_1(X), F_2(X), \cdots, F_n(X)) \tag{10}$$

$s.t.\ g_j(X) \leq 0 (j = 1, 2, \cdots, p)$

In the formula, $X$ is the strategy for a community discovery of signature network model and $g_j(X)$ is a community discovery constraint, and it can be defined as:

$$g^{te}(x | w, z^*) = \max_{1 \leq i \leq k} w_i | f_i(x) - Z_i^* |, s.t. x \in \Omega \tag{11}$$

In the formula, $z^* = (z_1^*, z_2^*, \cdots, z_k^*)$ is the reference point and $z_i^* = (\min f_i(x) | x \in \Omega)$.

## III. MULTI-OBJECTIVE OPTIMIZATION OF THE SIGNATURE NETWORK MODEL

### A. Particle Coding and Updating Rules

The particle position vector of the integer array is used here, and the binary code is conducted based on the character string. The adopted model is shown in Fig. 3.

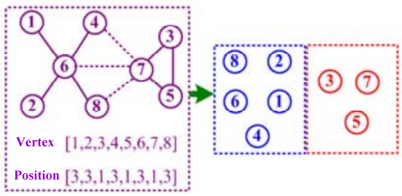



Figure 3. Particle representation model based on character string

Each particle represents a solution to optimize the problem. Traditional particle updating rule is that:

$$V_i \leftarrow V_i + c_1 r_1 (P_i - X_i) + c_2 r_2 (G - X_i) \tag{12}$$

$$X_{i+1} \leftarrow X_i + V_i \tag{13}$$

It can be seen from Fig. 3 that all the particle updating rules in the formula (11~12) are not suitable for the problem of signature network community discovery, so the new definition of the particle updating rule is that:

$$V_i \leftarrow \varphi(\omega V_i + c_1 r_1 (P_i \cap X_i) + c_2 r_2 (G \cap X_i)) \tag{14}$$

$$X_{i+1} \leftarrow X_i \cup V_i \tag{15}$$

In formula (13), "$\cap$"is the xor operation, and $\varphi(t)$ can be defined as :

$$\varphi(t) = \begin{cases} 1, \ if \ \ rand(0,1) \le 1/(1+e^{-t}) \\ 0, \ if \ \ rand(0,1) > 1/(1+e^{-t}) \end{cases} \tag{16}$$

In formula (13), the operator "$\cup$"can be defined as follows:

$$\begin{cases} X_i \cup V_i = X_i' = (x_{i1}', x_{i2}', \cdots, x_{in}') \\ x_{ij}' = x_{ij}, \ if \ \ v_i = 0 \\ x_{ij}' = \arg\max_r \sum_{k \in N_j} \delta(x_{ik}, r), \ if \ \ v_i = 1 \end{cases} \tag{17}$$

In the formula, $N_j$ is the neighborhood particle set of node $j$ ; if $a = b$ , $\delta(a,b) = 1$ , or $\delta(a,b) = 0$ . The operation steps are shown in Fig. 4.

According to Fig. 4, update the topology information particle state rule for signature network, the algorithm can be guaranteed to obtain a feasible solution to the community discovery.

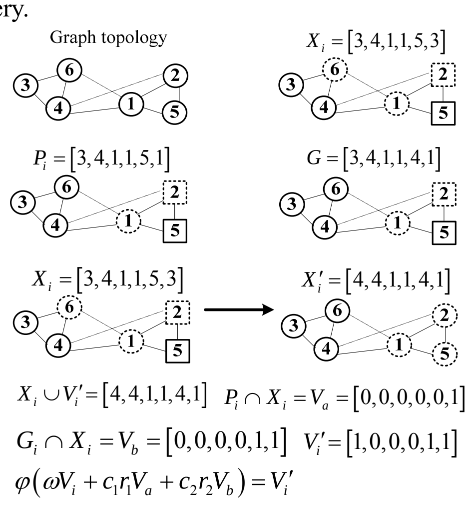



Figure 4. Particle updating rule

*B. Position Repair and Particle Replacement*

According to the coding method, 2 vectors in different positions may correspond to the same network community structure, as shown in Fig 5 in specific.

$$X_i = [1,2,3,3,4,1] \cap P_i = [1,2,3,3,4,1] = V_i = [0,0,0,0,0,0]$$

Repair Repair

$$X_i = [3,4,1,1,5,3] \cap P_i = [6,1,2,2,4,6] = V_i = [1,1,1,1,1,1]$$

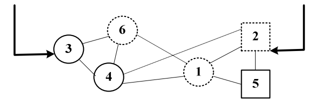


Figure 5. Community Structure of Position Vector

In Fig. 5, $X_i$ and $P_i$ respectively correspond to the best particle positions in the current and history. After decoding, $X_i$ and $P_i$ represent the same network community structure. According to formula (13~14), it can be obtained that the nonzero vectors $V_i$ and $X_i$ will change by time. To save time, design the location repairing algorithm, as shown in the pseudo code 1.

| Pseudo Code 1: Location Repairing Algorithm |
|---|
| 1. As for the given location vector $X$ , search for the individual element and stored in $U$ ;<br>2. for $i=0$ ; $i<U.size()$ ; $i++$<br>3. for $j=0$ ; $j<X.size()$ ; $j++$<br>3. if $X[j]==U[i]$ then $X[j]=i+1$ ;<br>4. endfor<br>5. endfor |

As shown in Fig. 5, after the location repairing, it can obtain the consistent zero vectors of $X_i$ and $P_i$ according to the formula (13). It does not need to calculate the new position vector, saving the computation time.

Replacement is intended to replace the original individual with a better offspring solution. This paper proposes a new updating strategy for sub-problem to keep the population diversity better. Supposed that $x_i^{t+1}$ is the new offspring solution, in the proposed updating strategy, only $T$ sub-problems get updated; the proposed updating strategy is shown in pseudo code 2.

| Pseudo Code 2: Updating Strategy for Sub-Problem |
|---|
| 1. Supposed that $B=\{1,2,\cdots,pop\}$ and $temp=0$ ;<br>2. for $j \in B$ do<br>3. if $g^{te}(x_i'\|\omega_j,z^*) \le g^{te}(x_j'\|\omega_j,z^*)$ then $temp=temp+1$ ;endif<br>4. if $temp \le T$ then<br>5. for $j \in B$ do<br>6. if $g^{te}(x_i'\|\omega_j,z^*) \le g^{te}(x_j'\|\omega_j,z^*)$ then $x_j'=x_i'$ , $F(x_j')=F(x_i')$ ;endif<br>7. else<br>8. for $j \in B$ && $g^{te}(x_i'\|\omega_j,z^*) \le g^{te}(x_j'\|\omega_j,z^*)$ do<br>9. Sort the Euclidean distance $F(x_i')$ of $F(x_j')$ and by descending;<br>10. Choose the first $T$ generations of solution and set $x_j'=x_i'$ and $F(x_j')=F(x_i')$ ;<br>11. endfor<br>12. endif |

### *C. Community Discovery of Signature Network Model Based on FS-UD-PSO*

**Process 1:** (quick sorting) conduct the initialization based on the non-dominance pair population, and carry out the initial individual sorting at the same time:

As for the population $P$ of the PSO algorithm, it has that all individuals adopt quick sorting according to the following operating procedure

①Firstly, supposed that $S_p = \phi$ and $n_p = 0$, the individual $p$ is a particle inside the population $P$ of the PSO algorithm, $S_p$ is the particle that is dominated by the particle $p$ inside the particle population, $n_p$ is the number of the dominated particle of individual $p$ inside the particle population;

②As for the particle $q$ inside the particle population $P$, if it meets the condition of $p \succ q$, it can be obtained that $S_p = S_p \cup \{q\}$; if it does not meet the relation of $q \succ p$, it can be obtained that $n_p = n_p + 1$.

③ $F1 = F1 \cup \{p\}$ ; Supposed that $n_p = 0$ , the particle level of particle $p$ is $p_{rank} = 1$; subjoin $p$ into the existing frontier Pareto, it can be obtained that $F1 = F1 \cup \{p\}$ ;

Execute the following steps circularly until it meets the preset termination condition $F_i = \phi$:

①Firstly, initialize the value of $Q = \phi$, aiming to store $F_i$ temporarily;

②As for all the particle $p$ inside the objective $F_i$ as well as all the particle $q$ inside the $S_p$ , operate as the following steps: set $n_q = n_q - 1$; if it meets $n_q = 0$ , which means that $q$ is only dominated by particle $p$, the sorting level of $q$ is $q_{rank} = i + 1$; supposed that $Q = Q \cup q$ at the same time;

③Supposed that $i = i + 1$ ;

④Supposed that $F_i = Q$ ; then No. $2 \sim n$ sorting frontier objective can be obtained in accordance with the sequence.

**Process 2:** (uniform density) in the non-dominance optimization steps of particle swarm population, the principle of retention of particles is to consider the particles with high adaptation degree first, followed by considering the particles with lower density; supposed that the current population contains $r$ sets of sub-objective $f_1$, $f_2$ … $f_r$ and that the area density of particle $i$ is $P[i]_{dis}$, then $P[i].m$ is the fitness of particle $i$ relative to sub-objective $m$ , so the uniform density indicator can be expressed as[12]:

$$P[i]_{dis} = \sum_{k=1}^{r} \left( P[i+1].f_k - P[i-1].f_k \right) \tag{18}$$

If the particle size of FS-UD-PSO is set to be $N$ , the maximum complexity of FS-UD-PSO optimization process is the quick sorting of $r$ sets of sub-function of the particle population; the complexity of the process is $O(rN \log N)$ ; the computational complexity of congestion is $O(rN)$ ; the complexity of the calculation process of the average density is $O(rN \log N)$ .

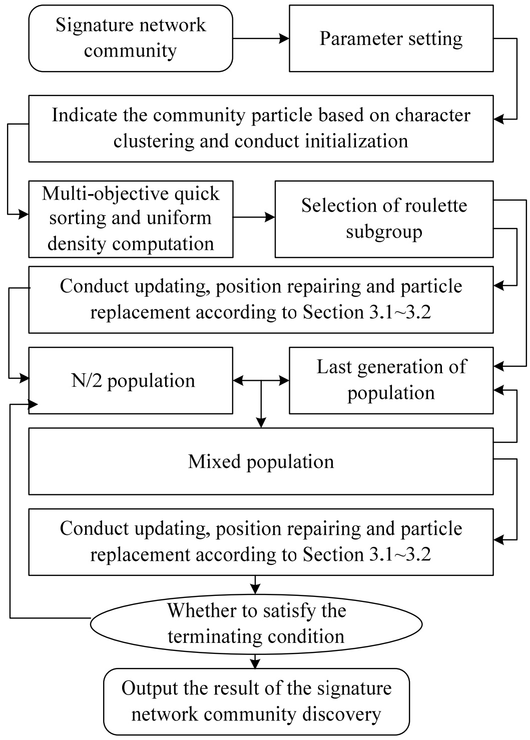



Figure 6. Process of FS-UD-PSO signature network community discovery

Process for signature network model of community discovery based on FS-UD-PSO algorithm is shown in Fig. 6 and the specific calculation processes are as follows:

**Step1:** Supposed that the population size of the initial particle of FS-UD-PSO algorithm is N, terminate algebra of particle evolution is $g_{\max}$ and meanwhile supposed that the particle population range is $[X_{\min}, X_{\max}]$ based on the uniform distribution method to initial of particle population x. Then evaluate the initial grade according to the index in Section 1.2 and conduct rapid dominated sorting and calculate the density of uniform region and suppose $i = 1$ at the same time.

Step2: Based on roulette method, obtain $N/2$ group individual particle in particle population *pop* to build parent particle $X_p$ and implement the operation of particle updating, location repair and particle displacement according to the contents of Section 3.1~3.2. The particle population scale can be obtained as $N/2$ after updating.

**Step3:** Blend the new particle population $X_{i+1}$ with previous generation particle population $X_{inter}$ , and conduct rapid sorting for the execution process 1 of T production particle population and conduct uniform density calculation of process 2 and conduct the new particle population *pop* containing $N$ groups of particles according to the calculation numerical value.

**Step4:** assumed that $i = i + 1$ , if meet the condition $i \le gen$ , then re-implement Step2; if meet the condition $i > gen$ , then continue to implement Step5.

**Step5:** Output the pareto optimal discovery strategy of signature network model community solved by FS-UD-PSO algorithm.

## IV. SIMULATION EXPERIMENT AND ANALYSIS

### A. *Experiment Settings*

Here conduct test for the benchmark signature network based on FS-UD-PSO algorithm and conduct algorithm coding through C++. The experiment hardware is i7-6800HQ, 2.8GHz, 8GB internal storage ddr3-1600GHz. Comparing algorithm selects MODPSO and MOEA-SN. The population scale $pop$ and the maximum algorithm iterations are set as 200, variation and crossover probability are set as 0.25 and 0.75 respectively, learning parameter is set as $c_1 = c_2 = 1.452$ and the inertial weight is set as $\omega = 0.725$. The adopted benchmark test set is shown as Table 1.

Table 1. Information statistics of signature network

| Network | Vertex number | Edge number | Positive edge number | Negative edge number |
|---|---|---|---|---|
| SPP | 11 | 44 | 16 | 26 |
| GGS | 15 | 57 | 28 | 24 |
| EGFR | 328 | 772 | 513 | 253 |
| Macrophage | 676 | 1417 | 936 | 473 |
| Yeast | 689 | 1073 | 859 | 221 |
| Ecoli | 1459 | 3208 | 1875 | 1329 |

### B. *Parameter Influence Experiment*

In updating operation, parameter $T$ will affect the algorithm performance of signature network community discovery, the following is about the test for its influence. The different value of parameter $T$ will result in the difference of Pareto frontier. Hypervolume index (HI) is adopted to evaluate the Pareto frontier effect.

$$HI\left(PS, y_{ref}\right) = L\left(\cup_{y \in PS}\left\{y' \middle| y \prec \middle| y' \prec y_{ref}\right\}\right) \tag{19}$$

Among them, $PS$ is the optimal solution set; $y_{ref} \in R^m$ shows all reference point while Pareto optimal solution is predominant, $m$ is the object quantity; $L$ represents the Lebesgue measure; $\prec$ represents dominance relation.

Normalize the $HI$ index, set the reference point $y_{ref}$ as 1.2, the change range of $T$ as $T = [1,3,5]$, the inertia weight as $\omega = 0.725$. The test result of benchmark signature network based on FS-UD-PSO algorithm is shown as Table 2.

According to the experiment statistics in Table 2, we can know that as the value of parameter $T$ increases, the HI index increases first and then decreases, the HI index when $T = 3$ is superior to the HI index when $T = 1$ and $T = 5$, and the HI index variance is relatively lower, showing that the algorithm stability is relatively better. The primary reason lies in that as the updating quantity of subproblem, if the value of $T$ is too big, it will waste a lot of computation time and it is not good for the reservation of algorithm evolution advantages and the diversity will greatly reduce. Meanwhile, with the increase of the value of parameter $T$, the computation complexity will increase correspondingly. However, if the value of $T$ is too small, it will not good for the improving of population evolution speed. Here, select $T = 3$ to conduct experiment according to the experiment result.

Table 2. Experiment result of parameter T influence

| Network | HI index | $T=1$ | $T=3$ | $T=5$ |
|---|---|---|---|---|
| SPP | Mean | 1.422 | 1.428 | 1.423 |
| | Std | 0.003 | 0.004 | 0.001 |
| GGS | Mean | 1.305 | 1.325 | 1.273 |
| | Std | 0.008 | 0.015 | 0.041 |
| EGFR | Mean | 1.134 | 1.143 | 1.069 |
| | Std | 0.032 | 0.042 | 0.098 |
| Macrophage | Mean | 1.213 | 1.237 | 1.187 |
| | Std | 0.027 | 0.019 | 0.069 |
| Yeast | Mean | 1.154 | 1.268 | 1.128 |
| | Std | 0.036 | 0.033 | 0.051 |
| Ecoli | Mean | 1.121 | 1.352 | 1.036 |
| | Std | 0.037 | 0.032 | 0.112 |

The setting method of inertia weight will affect the algorithm performance, select such three methods as inertia weight $\omega = 0.725$, $\omega \in [0.6, 0.8]$ random selection and nonlinear decreasing (quadratic curve decreasing), set $T$ =3. The result is shown as Table 3.

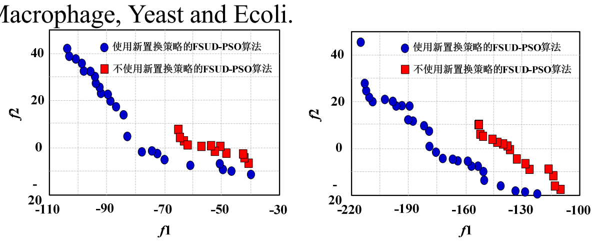
Table 3. Value comparison of inertia weight

| Network | HI index | $\omega = 0.725$ | $\omega \in [0.6, 0.8]$ | Quadratic curve decreasing |
|---|---|---|---|---|
| SPP | Mean | 1.428 | 1.438 | 1.452 |
| | Std | 0.004 | 0.009 | 0.008 |
| GGS | Mean | 1.325 | 1.342 | 1.386 |
| | Std | 0.015 | 0.012 | 0.014 |
| EGFR | Mean | 1.143 | 1.173 | 1.231 |
| | Std | 0.042 | 0.029 | 0.038 |
| Macrophage | Mean | 1.237 | 1.368 | 1.427 |
| | Std | 0.019 | 0.017 | 0.015 |
| Yeast | Mean | 1.268 | 1.239 | 1.316 |
| | Std | 0.033 | 0.028 | 0.032 |
| Ecoli | Mean | 1.352 | 1.386 | 1.453 |
| | Std | 0.032 | 0.029 | 0.031 |

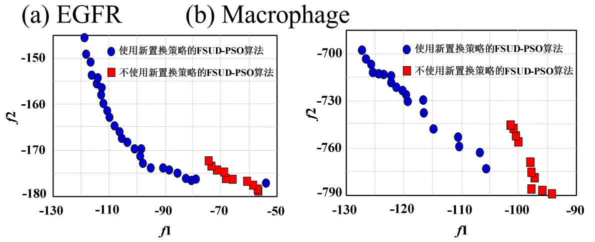
According to the comparison statistics in Table 3, we can know that the algorithm performance obtained by the inertia weight setting method of nonlinear decreasing (quadratic curve decreasing) is the optimal; the performance of $\omega \in [0.6, 0.8]$ random selection takes the next place. In the aspect of index stability, the stability of $\omega \in [0.6, 0.8]$ random selection is the optimal, but it has little difference with that of inertia weight setting method of nonlinear decreasing (quadratic curve decreasing).

### *C. Particle Displacement Experiment*

In order to check the effectiveness of the proposed strategy, here compare the FS-UD-PSO algorithm using new displacement strategy and the FS-UD-PSO algorithm without using new displacement strategy. The experiment result is shown as the Fig. 7. Because of the length, here just propose the comparison situation of Pareto optimal solution scheme based on four benchmark libraries of EGFR, Macrophage, Yeast and Ecoli.




(a) EGFR (b) Macrophage



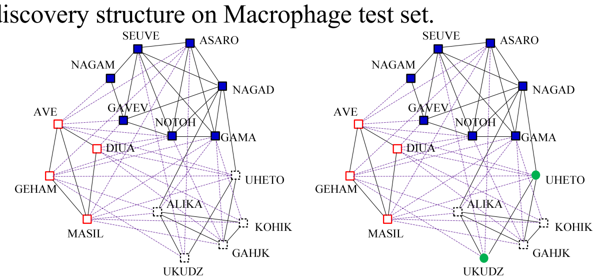

(c) Yeast (d) Ecoli

Figure 7. Experiment of displacement operation influence

The Fig. 7 shows the comparison situation between FS-UD-PSO algorithm using new displacement strategy and FS-UD-PSO algorithm without using new displacement strategy. According to the result in Fig. 7, we can know that the proposed new displacement strategy can effectively promote the discovery effect of signature network based on FS-UD-PSO algorithm, showing the effective of designed displacement strategy.

### *D. Performance of Community Discovery*

For each network, select the Pareto frontier with maximum signature module value solution as the proposed algorithm for final output. Fig. 8 shows the community discovery structure on Macrophage test set.




(a) Real community (b) Community discovery result in this paper

Figure 8. Macrophage test set

Fig. 8 shows the comparison between real community and community discovery result in this article on Macrophage test set. It shows that the original community discovery structure can be divided into 3 communities, while there are 4 communities in the discovery result in this paper. Therefore, the proposed algorithm can find other meaningful community structure.

The experiment result of 6 kinds of data set in Table 2 based on MODPSO, MOEA-SN and the proposed FS-UD-PSO algorithm is shown as Table 4. The comparison index selects the signature modularization $SQ$. According to the statistics in Table 4, the proposed FS-UD-PSO algorithm has higher $SQ$ value of signature modularization compared with other two algorithms, which shows that the FS-UD-PSO community discovery effect is superior to comparison algorithm. For example, in Macrophage test set, the $SQ$ index of FS-UD-PSO increases 7.1% and 8.6% respectively compared with the $SQ$ index of MODPSO and MOEA-SN algorithm.

Table 4. Index comparison of signature modularization

| Index | Algorithm | SPP | GGS | EGFR | Macrophage | Yeast | Ecoli |
|---|---|---|---|---|---|---|---|
| Cluster number | MODPSO | 53 | 68 | 94 | 86 | 102 | 123 |
| | MOEA-SN | 51 | 61 | 92 | 81 | 98 | 122 |
| | FS-UD-PSO | 43 | 54 | 83 | 72 | 87 | 124 |
| Signature modularization index | MODPSO | 0.2365 | 0.2594 | 0.2848 | 0.3027 | 0.5968 | 0.3975 |
| | MOEA-SN | 0.2257 | 0.2483 | 0.2731 | 0.2988 | 0.5876 | 0.3654 |
| | FS-UD-PSO | 0.2468 | 0.2691 | 0.2876 | 0.3244 | 0.6038 | 0.4031 |

## V. CONCLUSION

This paper proposes a community discovery algorithm of FS-UD-PSO signature network based on location repairing and particle replacement, improves the effect of signature network community discovery, gives multi-objective Pareto optimal target model of the signature network community on the basis of considering the data coupling and dependence, designs the position repairing and the particle replacement according to the characteristic of the signature network and designs multi-objective particle swarm optimization algorithm with quick sorting and uniform density to conduct signature network community discovery; the experimental results verify the effectiveness of the proposed community discovery of FS-UD-PSO signature network.

## REFERENCES


[1] Mukhopadhyay A, Maulik U, Bandyopadhyay S, et al. Survey of Multiobjective Evolutionary Algorithms for Data Mining: Part II[J]. IEEE Transactions on Evolutionary Computation, 2014, 18(1):20-35.

[2] Vialardi C, Bravo J, Shafti L, et al. Recommendation in Higher Education Using Data Mining Techniques.[J]. Journal of Animal Ecology, 2014, 84(2):326–336.

[3] Mccue C. Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis[J]. Data Mining & Predictive Analysis, 2015, 29(4):51–74.

[4] Fule P, Roddick J F. Detecting privacy and ethical sensitivity in data mining results[C]// Australasian Conference on Computer Science. Australian Computer Society, Inc. 2015:159--166.

[5] Haoudi A, Bensmail H. Bioinformatics and data mining in proteomics[J]. Expert Review of Proteomics, 2014, 3(3):333-343.

[6] Erlich Z, Gelbard R, Spiegler I. Evaluating A Positive Attribute Clustering Model for Data Mining[J]. Journal of Computer Information Systems, 2016, 43(3):100-108.

[7] Tsai C W, Lai C F, Chiang M C, et al. Data Mining for Internet of Things: A Survey[J]. IEEE Communications Surveys & Tutorials, 2014, 16(1):77-97.

[8] Ghosh S, Biswas S, Sarkar D, et al. A novel Neuro-fuzzy classification technique for data mining[J]. Egyptian Informatics Journal, 2014, 15(3):129-147.

[9] Wicker P, Christoph B. Analysis of problems using Data Mining techniques – findings from sports clubs in Germany[J]. European Journal for Sport & Society, 2016(2):131-140.

[10] Lausch A, Schmidt A, Tischendorf L, et al. Data mining and linked open data – New perspectives for data analysis in environmental research[J]. Ecological Modelling, 2015, 295:5-17.

[11] Zhang J, Li T, Chen H. Composite rough sets for dynamic data mining ☆[J]. Information Sciences An International Journal, 2014, 257(2):81-100.

[12] Adeniyi D A, Wei Z, Yongquan Y. Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method[J]. Applied Computing & Informatics, 2014, 36(1):90-108.

[13] Jinyu Hu and Zhiwei Gao. Distinction immune genes of hepatitis-induced heptatocellular carcinoma[J]. Bioinformatics, 2012, 28(24): 3191-3194.

[14] Liu, Y., Yang, J., Meng, Q., Lv, Z., Song, Z., & Gao, Z. (2016). Stereoscopic image quality assessment method based on binocular combination saliency model. Signal Processing, 125, 237-248.

[15] Jinyu Hu, Zhiwei Gao and Weisen Pan. Multiangle Social Network Recommendation Algorithms and Similarity Network Evaluation[J]. Journal of Applied Mathematics, 2013 (2013).

[16] Yishuang Geng, Jin Chen, Ruijun Fu, Guanqun Bao, Kaveh Pahlavan, Enlighten wearable physiological monitoring systems: On-body rf characteristics based human motion classification using a support vector machine, IEEE transactions on mobile computing, 1(1), 1-15, Apr. 2015

[17] Lv, Z., Halawani, A., Feng, S., Ur Réhman, S., & Li, H. (2015). Touch-less interactive augmented reality game on vision-based wearable device. Personal and Ubiquitous Computing, 19(3-4), 551-567.

[18] Jinyu Hu and Zhiwei Gao. Modules identification in gene positive networks of hepatocellular carcinoma using Pearson agglomerative method and Pearson cohesion coupling modularity[J]. Journal of Applied Mathematics, 2012 (2012).

[19] Jiang, D., Ying, X., Han, Y., & Lv, Z. (2016). Collaborative multi-hop routing in cognitive wireless networks. Wireless personal communications, 86(2), 901-923.

[20] Feng, L., & Lv, Z. (2016). Plane Surface Detection And Reconstruction Using Segment-Based Tensor Voting. Journal Of Visual Communication And Image Representation.

[21] Guo, H., Li, X., Wang, W., Lv, Z., Wu, C., & Xu, W. (2016). An Event-Driven Dynamic Updating Method For 3d Geo-Databases. Geo-Spatial Information Science, 1-8.

[22] Zhao, F., Li, G., Zhang, R., Du, J., Guo, C., Zhou, Y., & Lv, Z. (2015). Swarm-Based Intelligent Optimization Approach For Layout Problem. Multimedia Tools And Applications, 1-17.