Chapter 5

Community Detection using Particle Swarm Optimization

In chapter 3 and chapter 4, communities are identified in the networks having single connections among nodes. In this chapter, multiple featured networks are considered for detecting communities. An evolutionary technique is studied and utilized for efficient identification of communities by optimizing community related objective function.

5.1 Introduction

Real-world systems are complex as those cover wide range of aspects such as multiple relationships [9, 114, 164, 192], so notion of multiple featured network is developed. Connections in multiple featured networks are much more complicated than the networks where single connections exist among nodes. A multiple featured network contains several networks with same set of nodes. Each of those networks is an independent network involving specific feature. Thus, multiple connections are present between two nodes in multiple featured networks, and as a result identification of communities becomes more difficult. Often, community detection problem is translated into a combinatorial optimization problem with the goal of optimizing an objective function. However, community detection in the networks having single connections among nodes is a NP-Hard [54, 156]. Moreover, Brandes et al. [20] shown modularity maximization in those networks is NP-Complete problem. Heuristic approaches have been incorporated to deal with the complexity of community detection problem in such networks, particularly nature-inspired evolutionary computation techniques (see [26] for the reviews).

Nature-inspired heuristic techniques can be applied to any network, regardless of whether it possesses a community structure or not. Simply requires to define appropriate objective function that is to be optimized. Particle Swarm Optimization (PSO) is widely used to optimize the objective function related to community detection [6, 41, 181]. In this chapter, a cognitive avoidance mechanism is introduced to improve the performance of PSO. A suitable objective function is defined for identifying communities in multiple featured networks. This objective function is optimized using the improved PSO with cognitive avoidance mechanism.

5.2 PSO with Cognitive Avoidance Mechanism

Main inspiration behind the PSO algorithm [47, 99] is social behavior of swarms such as birds flocking, fish schooling. PSO maintains a population of particles referred as swarm. Individual solution is considered as particles in the swarm. Each particle interacts with one another and simultaneously learns from its own experience. As PSO learns from its neighbor, particles with higher fitness put comparatively greater impact on swarm behavior. Each particle is associated with a velocity and a position in the solution domain. Every particle maintains their best solution experienced so far i.e. the local best (pbest) and best solution experienced by neighbor i.e. the global best (gbest). Positions and velocities are updated in accordance with the current pbest and gbest.

Position, velocity, pbest and gbest vector of i^{th} particle at t^{th} iteration in d dimension can be represented respectively as shown below:

$$X_i(t) = (x_{i1}, x_{i2}, x_{i3}, ...x_{id})$$

$$V_i(t) = (v_{i1}, v_{i2}, v_{i3}, ...v_{id})$$

$$P_i(t) = (p_{i1}, p_{i2}, p_{i3}, ...p_{id})$$

$$G_i(t) = (g_{i1}, g_{i2}, g_{i3}, ...g_{id})$$

Velocity and position of i^{th} particle in j^{th} dimension for next iteration are updated with the following two equations:

$$V_{ij}(t+1) = V_{ij}(t) + C_1 \times R_1 \times (P_{ij}(t) - X_{ij}(t)) + C_2 \times R_2 \times (G_{ij}(t) - X_{ij}(t)) \quad (5.1)$$

$$X_{ij}(t+1) = X_{ij}(t) + V_{ij}(t+1) \quad (5.2)$$

where, R_1 and R_2 are uniformly distributed random numbers in range [0, 1]. C_1 and C_2 are the positive constants in range (0, 2], known as acceleration coefficients. C_1 controls particle's movement towards local best and C_2 controls particle's movement towards global best. The term $C_1 \times R_1 \times (P_{ij}(t) - X_{ij}(t))$ is associated with particle's cognition of its own best solution. The term $C_2 \times R_2 \times (G_{ij}(t) - X_{ij}(t))$ is associated with particle's collaborative interaction with its neighbors. These two terms are related with particle's acceleration (rate of change in velocity) so these are often known as cognitive acceleration and social acceleration respectively.

Despite the fact that, pbest and gbest motivates particles to move towards optimal solution, interference of random values may cause particles to move to some unfruitful positions. These unfruitful movements may degrade algorithms overall performance. Any wrong movement of a particle may lead to diverse from optimal solution, which may result delay in convergence. Effect of such unfruitful movement propagates in successive iteration resulting extra iterations to reach the destination or may diverted to completely different direction. Any misguidance may slow down the overall convergence of algorithm towards the optimal solution. Awareness of such pitfalls may improve overall performance of PSO by avoiding them. Therefore, very similar mechanism that pbest and gbest does to a particle to attract towards themselves is used to develop cognitive avoidance mechanism by making particles aware of such unfruitful solutions and avoid movement towards those.

Cognitive Avoidance Mechanism: Each particle maintains its worst value that it has attained so far along with the pbest and gbest. With this known worst value particles tries to avoid further movement towards it, having the sense that solutions nearby the worst one may not be suitable. Particle's own known worst solution so far (pworst) pushes particles backward so that they can never be trapped again into it. This avoidance mechanism also reduces movement of particles towards other bad solutions around the pworst. This is a kind of inverse greedy approach where particles are distracted instead of attracting as pbest and gbest does. To define such avoidance scheme, a new component referred as cognitive avoidance component is added to the existing velocity equation of PSO. Current pworst vector of i^{th} particle is represented as $W_i(t) = (w_{i1}, w_{i2}, w_{i3}, ...w_{id})$ where, d is the dimension of particle. Velocity equation is redefined as follows:

$$V_i(t+1) = \omega * V_i(t) + C_1 * R_1 * (P_i(t) - X_i(t)) + C_2 * R_2 * (G_i(t) - X_i(t)) - C_3 * R_3 * (W_i(t) - X_i(t)) \quad (5.3)$$

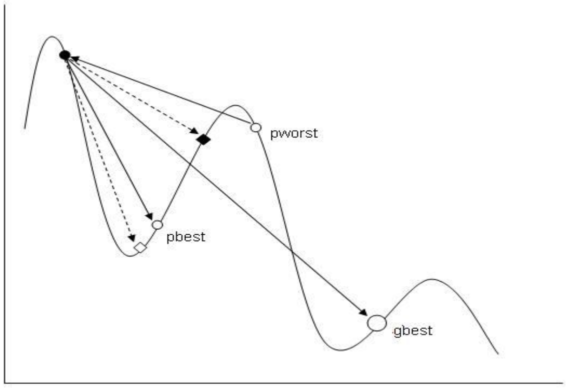


Figure 5.1: Effect of cognitive avoidance in PSO

Fourth component in Equation 5.3 represents cognitive avoidance and is considered as negative since it represents distraction, which is opposite to cognitive acceleration and social acceleration. To control the effect of this avoidance to a particle, cognitive avoidance coefficient C_3 is used along with randomness R_3 in range [0, 1]. Position equation remains unaltered as in Equation 5.2. This new addition to the existing PSO is referred as Particle Swarm Optimization with Cognitive Avoidance (PSOCA). Effect of the newly added component is shown with an example in Figure 5.1. Black circle represents current position of a particle. Black diamond represents next position of the particle guided by pbest and gbest only. White diamond represents next position influenced by cognitive avoidance component. Proposed approach avoids movement towards worst solution and pushes final solution towards either pbest or gbest. In this case it moves towards pbest.

A population of particles is initialized with randomly generated positions and velocities. Fitness of each particle is evaluated with user defined objective function. In each generation, velocity of a particle is updated with Equation 5.3 and next position of particle is evaluated with Equation 5.2. Any particle Updates its current pbest, pworst and gbest. Generally, velocity of particles are controlled with predefined velocities. If any particle gains larger velocity than the predefined velocity, modulus of the velocity is considered for updating positions. Any predefined velocity limits for particles is not considered. Particles are allowed to move with any finite velocity. If a particle moves outside the search space then that movement is controlled with defined limits of each dimensions.

Although, strategically overcome the problem of unfruitful moves, but the problem of misguidance still remains in PSOCA. As the method incorporates probabilistic move by avoiding unfruitful moves, which may again misguide particles in some cases where optimal value is nearby the present pworst value. Particle may wrongly avoid move towards the optimal solution and may never reach the optimal solution. Here, there has to be some mechanism so that such misinterpretation can be compensated in successive iterations. To overcome such situation a varying cognitive avoidance coefficient (C_3) is introduced to PSOCA. A very similar mechanism used for varying weight in PSO-TVIW [180] is utilized here for varying C_3. This extension of PSOCA is referred as PSO with Time Varying Cognitive Avoidance (PSO-CATV). C_3 is updated in each iteration as follows:

$$C_3 = C_{3min} + (C_{3max} - C_{3min})\left(\frac{mIter - Iter}{mIter}\right) \tag{5.4}$$

where, *mIter* is the maximum number of iterations and *Iter* is the current iteration number. It is clear from the Equation 5.4 that the value of C_3 gradually decreases with successive iterations. Hence, effect of cognitive avoidance reduces as iteration increases. Thus, even if situation arises where optimal solution is nearby the pworst, particle can reach the optimal solution as effect of misinterpretation reduces.

5.3 Community Detection using PSOCA

In this section, representation of multiple featured network is demonstrated with a suitable example. Explained representation of particles in the context of multiple featured

Persons	a_1	a_2	a_3	a_4
P_1	1	0	1	1
P_2	1	1	1	1
P_3	0	1	1	1
P_4	1	0	0	1
P_5	1	1	0	0

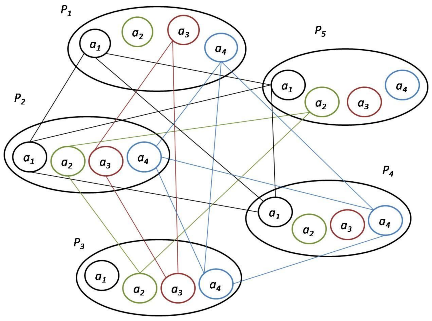


Figure 5.2: Multiple featured network representation of instances with attributes. A person is connected with respect to an attribute to other if both have entries 1 in the table.

network. An objective function related to community detection is designed for optimizing it using PSOCA. A generic algorithm is designed, which is also applicable for other variants of PSO to detect communities.

5.3.1 Multiple Featured Network Representation

Social life is full of choices. Everyone has different interests, likings and dis-likings. For instance, a person may like Cricket and at the same he may dislike Hockey, or he may like to listen Hindi music over English and so on. Hence people can be attributed to various things based on their interests. For this particular example, there are four attributes/features Cricket, Hockey, Hindi music and English music. Two possibilities are associated with each of the features, either it is liked or disliked by an individual. In real life situations there may be several instances of this kind. The liking and dis-liking is represented with the numerical values 1 and 0 respectively. The information is transformed into a boolean table where each row or instance represents a person's list of likings and dislikings and such instances can be represented with multiple featured network as shown in Figure 5.2. Aim is to identify network of people or communities having the most overlapping likes or dislikes. Such groups may consist mostly of people liking hindi music and cricket or a group which is comprised of hockey and cricket liking people and so on.

Choices vary from person to person since everyone has his own likes and dislikes. These likes and dislikes may match with other persons likes and dislikes completely, partially or may not match at all. In this situation it becomes an interesting task to identify groups or communities of similar people on the basis of some criteria. Considering this fact, various choices of a person are represented as a vector in which the dimensions denote the features. The entry corresponding to a feature is 1 if the instance is associated with the feature otherwise 0. For an example, say there are two persons. Suppose, skills associated with them are dramatics and debating (these skills are features). Say the first person is good at dramatics and debating whereas the second has only good debating skills then this scenario is represented as two instances (1,1) and (0,1) respectively.

5.3.2 Particle Representation and Objective Function

The particles are represented in terms of instances. Given set of n instances containing m features, which have to be grouped into k communities. A particle is represented with k centroids for each of the k communities. Thus, each particle in the population represents different grouping of k communities i.e. a community structure. To determine fitness of a particle, Davies-Bouldin index [39] is considered. This index measures quality of a single community. Since a particle consists of k communities, Davies-Bouldin index of all communities are averaged to obtain fitness of the particle.

Let $P_x = \{A_1, A_2, ..., A_k\}$ be a particle that represents a community structure having k communities, where $A_i = \{a_{i1}, a_{i2}, ..., a_{im}\}$ is any instance vector of all feature values which has been chosen as the centroid for the i^{th} community. Let $A_i = \{a_{i1}, a_{i2}, ..., a_{im}\}$ and $A_j = \{a_{j1}, a_{j2}, ..., a_{jm}\}$ be the i^{th} and j^{th} instances respectively. For measuring similarity between two instances, Euclidean distance is considered. The Euclidean distance between

Algorithm 5.1: ComputeFitness(P_x)

```
Initialize centroids with each dimension of $P_x$
$sum \leftarrow 0$
for each centroid $A_j$ in $P_x$ do
    $S_j \leftarrow$Add Euclidean distances from all allocated instances $A_i$ to $A_j$
    $S_j \leftarrow S_j/$no. of allocated instances
    $sum \leftarrow sum + S_j$
end for
$sum \leftarrow sum/$no. of communities
return $sum$
```

two instances is computed as follows:

$$E_{ij} = \sqrt[2]{(a_{i1}-a_{j1})^2+(a_{i2}-a_{j2})^2+...,(a_{im}-a_{jm})^2}. \tag{5.5}$$

Consider that the community corresponding to the y^{th} dimension of any particle P_x as $C_y = \{A'_1, A'_2, ..., A'_t\}$, the centroid of which is represented in the particle as A_y. This community comprises t instances. The Davies-Bouldin index of the community is obtained as follows:

$$B_y = \frac{E_{y1}+E_{y2}+...+E_{yt}}{t}. \tag{5.6}$$

The Equation 5.6 gives individual Davies-Bouldin index of all communities corresponding to the centroids present in each dimension of a particle. To compute fitness of the particle P_x, objective function is defined as follows:

$$f(P_x) = \frac{1}{k}\sum_{y=1}^{k} B_y. \tag{5.7}$$

The pseudo code for fitness computation of each particle P_x is presented in Algorithm 5.1.

Algorithm 5.2: FindCommunity(*MaxGen*,*NumPar*,*N*,*K*)

```
// MaxGen denotes the maximum number of generations, NumPar denotes the number of particles to work with, N denotes the number of instances, K represents the number of communities to identify
Initialize particles of population with randomly selected instances.
$generation \leftarrow 0$
repeat
    $generation \leftarrow generation + 1$
    for each particle $P_x$, where $x = 1,2,3,...,NumPar$ do
        for each instance $A_i$, where $i = 1,2,3,...,N$ do
            Measure Euclidean distances from all centroid $A_j$ using Equation 5.5
            Assign $A_i$ to centroid $A_j$ having minimum distance
        end for
        $Fitness[P_x] = ComputeFitness(P_x)$
    end for
    Update personal best of each particle $P_x$
    Update global best of population
    Update position of each particle $P_x$ using position equation of PSO
    Update velocity of each particle $P_x$ using velocity equation of PSO
until $generation \leq MaxGen$
```

5.3.3 Proposed Algorithm

The algorithm follows similar approach as k-means algorithm in background for assigning each instances to a particular community. Unlike k-means, re-evaluation of centroids are not performed iteratively. Instead, centroids are determined during fitness evaluation of particles. For measuring similarity between two instances Euclidean distance is considered. PSOCA is used to detect communities. In each generation, the position and velocity of particles are updated based on the “gbest”, “pbest” and “pworst” of particles. An upper limit to the number of generations is specified in order to terminate the execution and chose the particle with best fitness in the population as the final solution. Pseudo code of proposed algorithm is presented in Algorithm 5.2.

5.4 Empirical Analysis

5.4.1 Experimental Setup for PSOCA and PSO-CATV

Benchmark Functions: Two sets of benchmark functions are considered for analyzing performance of improved versions of PSO with cognitive avoidance mechanism. First set comprises very popular 4 benchmark functions that were used for performance evaluation. Among these Rosenbrock's function is uni-modal, whereas Ackley's, Rastrigin's and Schwefel's functions are multi-modal function. All functions have global optimal solution at or near the origin except Schwefel's function, which have global optimal solution at the edges of the solution domain. Function definitions and their initial ranges are shown in Table 5.1 and 5.2 respectively.

Second set comprises 25 benchmark functions that were presented in CEC 2005 special session of real-parameter optimization [187].

- 5 unimodal functions
    - f1: Shifted Sphere Function.
    - f2: Shifted Schwefel's Problem 1.2.
    - f3: Shifted Rotated High Conditioned Elliptic Function.
    - f4: Shifted Schwefel's Problem 1.2 with Noise in Fitness.
    - f5: Schwefel's Problem 2.6 with Global Optimum on Bounds.
- 20 multimodal functions
    - 7 basic functions
        - f6: Shifted Rosenbrock's Function.
        - f7: Shifted Rotated Griewank Function without Bounds.

    - f8: Shifted Rotated Ackley's Function with Global Optimum on Bounds.
    - f9: Shifted Rastrigin's Function.
    - f10: Shifted Rotated Rastrigin's Function.
    - f11: Shifted Rotated Weierstrass Function.
    - f12: Schwefel's problem 2.13.
  - 2 expanded functions
    - f13: Expanded Extended Griewank's plus Rosenbrock's (Ef8f2)
    - f14: Shifted Rotated Expanded Scaffers F6.
  - 11 Hybrid Composition Functions
    - f15: Hybrid Composition Function 1
    - f16: Rotated Hybrid Composition Function 1
    - f17: Rotated Hybrid Composition Function 1 with Noise in Fitness
    - f18: Rotated Hybrid Composition Function 2
    - f19: Rotated Hybrid Composition Function 2 with a Narrow Basin for the Global Optimum
    - f20: Rotated Hybrid Composition Function 2 with the Global Optimum on the Bounds
    - f21: Rotated Hybrid Composition Function 3
    - f22: Rotated Hybrid Composition Function 3 with High Condition Number Matrix
    - f23: Non-Continuous Rotated Hybrid Composition Function 3
    - f24: Rotated Hybrid Composition Function 4
    - f25: Rotated Hybrid Composition Function 4 without Bounds

Optima of all the functions has been displaced from the origin or from the previous position to ensure that optimal solutions can never be obtained in center of the domain.

Table 5.1: Benchmark functions

Function	Definition
Rastrigin	$f(x) = \sum_{i=1}^{n} \left[x_i^2 - 10\cos\left(2\pi x_i\right) + 10\right]$
Ackley	$f(x) = (20+e) - 20\exp\left[-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}\right]$ $-\exp\left[\frac{1}{n}\sum_{i=1}^{n}\cos\left(2\pi x_i\right)\right]$
Rosenbrock	$f(x) = \sum_{i=1}^{n-1} \left[100\left(x_{i+1} - x_i^2\right)^2 + \left(x_i - 1\right)^2\right]$
Schwefel	$f(x) = \sum_{i=1}^{n} \left(x_i \sin\sqrt{\lvert x_i \rvert}\right)$

This displacement mechanism has made it difficult for the algorithms that have central tendency to attain optimal solutions.

Algorithms and Parameters: Since cognitive avoidance approach is added to PSO to improve overall performance, so proposed PSOCA is compared with standard PSO only. However, improved version of PSOCA i.e. PSO-CATV is compared with other two variants of PSO, PSO-TVIW [180] and PSO-TVAC [158]. Along with these two variants of PSO, another state-of-the-art competitor jDE [22] is also considered to compare performance of PSO-CATV. This is an extension to the Differential Evolution approach, which can adapt parameters CR and F, suitable to corresponding situation.

Acceleration coefficients of PSOCA C_1, C_2 and newly added C_3 are kept as constant values 0.6, 1.5 and 0.4 respectively. However for PSO-CATV, acceleration coefficients C_1 and C_2 are kept same constant value as PSOCA, but the newly added C_3 varied linearly from $C_{3max} = 2.0$ to $C_{3min} = 0.25$. For PSO-TIVW, the value of ω updated with $\omega_{min} = 0.4$ and $\omega_{max} = 0.9$ [180]. For PSO-TVAC, value of C_1 varied from 2.5 to 0.5, C_2 from 0.5 to 2.5 and ω from 0.9 to 0.4 [158]. The jDE algorithm does not need any parameter setting as its parameters CR and F are adapted during execution.

Table 5.2: Initial range and Optima

Function	Range	Optimal solution
Rastrigin	$[-5.12, 5.12]$	$f(x^*) = 0$
Ackley	$[-15, 30]$	$f(x^*) = 0$
Rosenbrock	$[-2.048, 2.048]$	$f(x^*) = 0$
Schwefel	$[-500, 500]$	$f(x^*) = -n \times 418.9829$

5.4.2 Analysis of Solution Quality

To analyze solution quality of PSOCA, first set of benchmark functions are tested with dimensions 10, 20 and 30. Population size is considered for this case is 40 and maximum iteration limit as 1000 for each run to made comparisons more precise. Second set consisting 25 CEC 2005 benchmark functions is divided in two subsets. Subset 1 includes first 14 functions and subset 2 includes remaining 11 hybridized functions. The functions of subset 1 is considered to analyze solution quality of PSO-CATV, while subset 2 is considered for convergence analysis. Dimensions are 50, 60 and 70 are considered. As increase in dimensionality increases complexity of the problem, so population size 100 is considered. Also considered different maximum iteration limit depending on there complexity level. For dimensions 50, 60 and 70, maximum iteration limit are considered as 3000, 4000 and 5000 respectively. Each function and corresponding considered dimension are executed over 50 trials to present performance metrics.

Results of PSOCA and PSO on the first set of benchmark functions, performance metrics of optimal solutions over 50 trials are presented as in Table 5.3. For Rastrigin's function in all dimensions PSOCA performs better than PSO in terms of both mean and standard deviation. For Ackley's function in dimensions 10 and 20 shows almost similar results, but in 30 PSOCA performs better than PSO. In dimension 10 both reaches optimal value. For Rosenbrock's function in dimension 10 performance of PSOCA is poor. In dimension 20, PSOCA performs very well. In dimension 30 shows little improvement in performance

Table 5.3: Comparison of PSO and PSOCA

Function	**Dimension**	**Measures**	**PSO**	**PSOCA**
Rastrigin	10	Mean	6.268238	5.890155
		Std. Deviation	3.205177	2.725471
	20	Mean	35.679164	30.087520
		Std. Deviation	11.133938	7.594830
	30	Mean	83.605148	80.671492
		Std. Deviation	19.746506	6.351290
Ackley	10	Mean	0.000000	0.000000
		Std. Deviation	0.000000	0.000000
	20	Mean	0.701055	0.511282
		Std. Deviation	0.821616	0.790906
	30	Mean	2.379818	2.202456
		Std. Deviation	0.878856	0.045834
Rosenbrock	10	Mean	0.495389	0.587538
		Std. Deviation	1.304553	1.397056
	20	Mean	12.151854	9.016740
		Std. Deviation	9.862681	3.645567
	30	Mean	35.199659	33.586877
		Std. Deviation	23.205804	23.058436
Schwefel	10	Mean	-3726.731438	-3733.443933
		Std. Deviation	186.940914	175.685358
	20	Mean	-6312.808570	-6388.208710
		Std. Deviation	393.573309	386.027770
	30	Mean	-8259.610620	-8280.060558
		Std. Deviation	584.277113	507.231340

than PSO. From these experiments it is clear that introduction of cognitive avoidance to PSO not only improves results but also gives benefit for higher dimensional problems. Performance of PSOCA seems better with increment of dimensionality.

Results of PSO-CATV and other competitor on first seven of subset 1 of CEC 2005 special session benchmark functions with the performance metrics of optimal solutions over 50 trials are presented as in Table 5.4. It is clear from the results presented in Table 5.4 that PSO-CATV outperforms over PSO-TVIW, PSO-TVAC and jDE in functions f1 and f2 for all dimensions. In f3, PSO-CATV shows better result than PSO-TVIW and PSO-TVAC,

Table 5.4: Comparison of PSO-CATV with PSO-TVIW, PSO-TVAC and jDE part I

f(x)	Dim	Measures	PSO-TVIW	PSO-TVAC	PSO-CATV	jDE
f1	50	Mean	3.902552	4.908112	0.000000	0.000000
		SD	12.940261	8.163804	0.000000	0.000000
	60	Mean	2.196993	8.109815	0.000000	0.000000
		SD	3.415574	16.602715	0.000000	0.000000
	70	Mean	4.089994	7.808844	0.000000	0.000000
		SD	11.795045	16.735104	0.000000	0.000000
f2	50	Mean	130.631028	70.339226	5.745377	11.48375
		SD	41.359146	31.479238	3.555487	4.28824
	60	Mean	329.163198	162.796134	17.549255	183.68
		SD	89.857498	59.878654	10.952076	47.192
	70	Mean	641.125525	413.048155	59.965942	132.897
		SD	153.346504	153.109885	39.948230	54.55045
f3	50	Mean	10808432.435	11897240.811	3045488.056	1570000.00
		SD	6285139.025	7327036.636	1146567.177	42426.40687
	60	Mean	14087070.935	11252155.437	4242028.077	4355950
		SD	6878314.411	7192879.592	1550749.991	888621.091
	70	Mean	16878072.985	14521790.405	5933886.299	5425700.00
		SD	10255231.182	8592850.856	2364138.274	1423123.108
f4	50	Mean	3152.656009	7947.115621	3566.457645	6107.6
		SD	1128.932254	2596.575130	13858.383438	431.9008219
	60	Mean	6632.142723	14127.583262	8029.763677	18038
		SD	1842.963164	4381.406693	3078.674737	9916.465499
	70	Mean	12292.978720	21543.842494	14205.333739	25808
		SD	3372.047619	4443.916660	4162.635158	4449.1158
f5	50	Mean	7995.021248	8232.194651	7328.712549	5722.7
		SD	1073.176077	1444.619689	1167.819993	771.73634
	60	Mean	10263.928239	10644.165517	10542.227027	12593.6
		SD	1878.519310	1510.495959	1925.873026	4499.179
	70	Mean	13694.404809	13451.695247	11685.651189	13239.5
		SD	2318.435464	2431.814152	1760.678965	178.89801
f6	50	Mean	212.248266	163.344817	96.338635	297.845
		SD	253.610211	126.706071	89.031715	137.89289
	60	Mean	108.252552	122.628378	73.948984	161.995
		SD	62.138997	43.533576	35.277724	4.3345
	70	Mean	135.399165	187.695732	110.006910	173.225
		SD	60.001736	114.768624	56.843038	18.6039
f7	50	Mean	0.995867	0.994219	0.008458	6195.3
		SD	0.004578	0.009273	0.012966	0.0000
	60	Mean	0.996477	0.999869	0.011748	7230
		SD	0.003516	0.012295	0.011688	0.0000
	70	Mean	0.998251	1.012891	0.008993	8675.8
		SD	0.002023	0.031169	0.007952	0.0000

Table 5.5: Comparison of PSO-CATV with PSO-TVIW, PSO-TVAC and jDE part II

f(x)	**Dim**	**Measures**	**PSO-TVIW**	**PSO-TVAC**	**PSO-CATV**	**jDE**
f8	50	Mean	21.115114	21.095453	21.119545	21.1465
		SD	0.048413	0.056330	0.042264	0.031819
	60	Mean	21.163687	21.165886	21.191519	21.1965
		SD	0.043717	0.044411	0.042322	0.0261
	70	Mean	21.200285	21.198274	21.231040	21.224
		SD	0.048999	0.040314	0.048431	0.04808
f9	50	Mean	291.899708	287.740901	186.063673	0.0000
		SD	28.896009	26.897532	209.810298	0.0000
	60	Mean	367.496346	360.830206	684.778702	0.0000
		SD	35.602923	33.208458	527.462884	0.0000
	70	Mean	430.257996	434.277740	1236.117304	0.0000
		SD	37.392734	33.017474	455.333121	0.0000
f10	50	Mean	508.576562	490.925751	469.395445	113.425
		SD	241.029088	66.944758	76.089387	7.03571
	60	Mean	1251.609138	605.504681	557.828100	134.815
		SD	619.303092	67.921327	82.476083	7.7428
	70	Mean	1900.690573	706.790286	683.369855	188.05
		SD	573.614498	85.963727	91.500468	5.62856
f11	50	Mean	84.356621	34.680733	33.995479	52.0355
		SD	6.656662	6.215943	5.392729	14.68024
	60	Mean	104.211198	45.519844	42.041221	60.6085
		SD	2.621380	6.917720	6.358932	20.4007
	70	Mean	122.432169	53.992128	51.324729	72.847
		SD	8.117650	7.914189	3.887204	27.08218
f12	50	Mean	64292.926	64548.556	3435963.583	36695
		SD	29754.286	34908.068	4708418.529	37611.01
	60	Mean	78812.533	87337.496	12242957.869	49456.5
		SD	33183.514	48051.435	5935467.345	1642.609
	70	Mean	127746.286	123025.408	18825897.822	50784.5
		SD	53646.227	56343.148	5704889.387	26141.0305
f13	50	Mean	243.747263	33.841916	10.225835	11.293
		SD	951.548735	4.519735	0.02174919	0.0028284
	60	Mean	2212.361500	42.905302	46.441168	16.467
		SD	3131.156487	7.183050	6.974479	0.32526
	70	Mean	6565.684216	52.922639	59.530015	20.859
		SD	3543.521298	6.606150	8.653405	2.070408
f14	50	Mean	21.83099	21.681058	21.757713	22.6245
		SD	0.529896	0.626972	0.444706	0.099702
	60	Mean	26.662176	26.813936	26.210791	27.3485
		SD	0.559658	0.589835	0.700757	0.089802
	70	Mean	31.400545	31.008263	31.735995	32.335
		SD	0.516103	0.875929	0.462124	0.1385

and shows very similar results to jDE. In f4, for all dimensions PSO-CATV outperforms jDE and PSO-TVAC, but fall behind PSO-TVIW. In f5, for dimensions 50 and 60 performance of PSO-CATV is almost similar to both the competitor, but for dimension 70 PSO-CATV shows drastic improvement. However, for dimension 50 performance of jDE is better than PSO-CATV and it experiences better in result than jDE on dimensions 60 and 70. Again in functions f6 and f7, PSO-CATV outperforms over PSO-TVIW and PSO-TVAC for all dimensions. However, jDE performs better than PSO-CATV in f6. Performance of PSO-CATV again seems better than jDE and noted that jDE is getting stick to some specific values over all 50 trials. Which indicates that jDE might have fall into that local optima and unable to come out of that optima.

Results on rest of the functions of subset 1 i.e. f8-f14 are presented in Table 5.5. In f8, performance of PSO-CATV is very similar performance as jDE, PSO-TVIW and PSO-TVAC. In f9, performance of PSO-CATV is poor. In f10, PSO-CATV outperforms in almost all dimensions than PSO-TVIW and PSO-TVAC, but poorer than jDE. Again in f11, PSO-CATV outperforms all three competitors in almost all dimensions. In f12, PSO-CATV shows very poor performance. However, all three competitors also show poor result in this function, though comparatively better than PSO-CATV. In f13, PSO-CATV outperforms PSO-TVIW but, falls behind PSO-TVAC and jDE. In f14, performance of PSO-CATV is almost similar but comparatively better than all three competitors. Overall performance of PSO-CATV is better than all three PSO-TVIW, PSO-TVAC and jDE in all functions except f9 and f12.

5.4.3 Analysis of Convergence

For convergence analysis considered subset 2 of CEC 2005 special session benchmark functions that includes f15 to f25 of CEC 2005 benchmark functions. As subset 2 consist

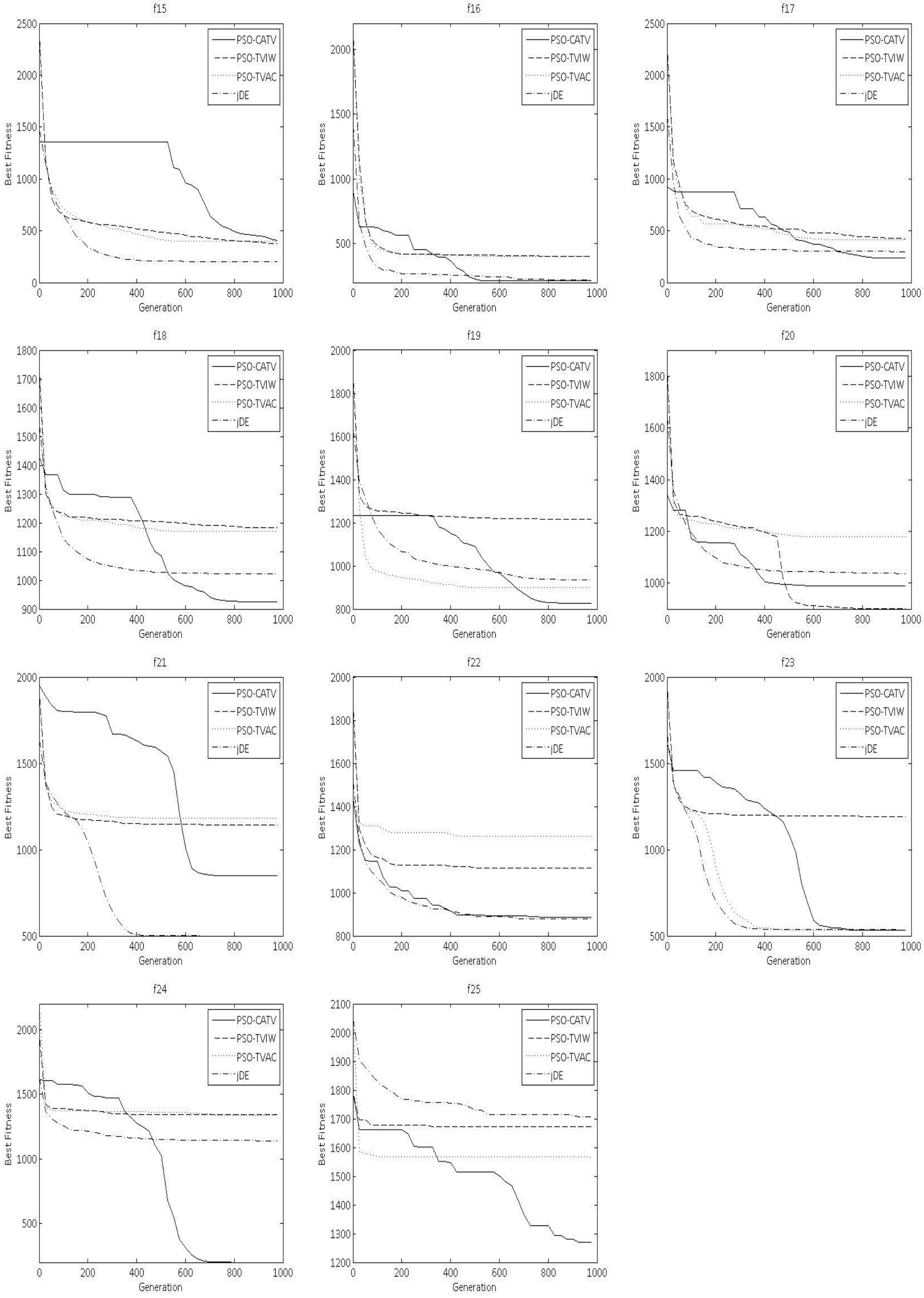


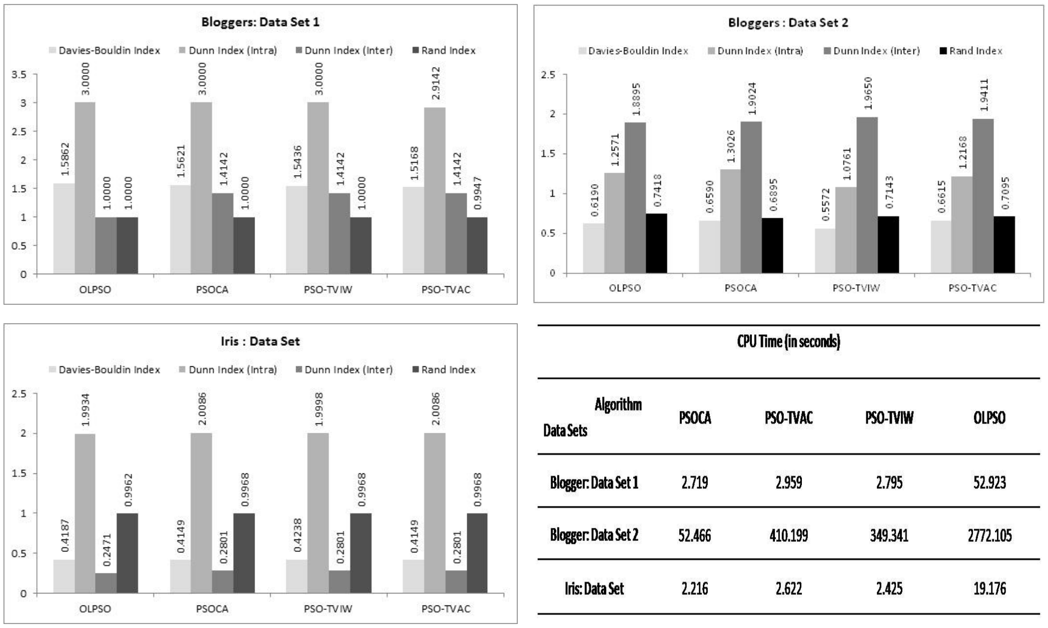
Figure 5.3: Convergence comparison of PSO-CATV with PSO-TVIW, PSO-TVAC and jDE.

11 functions, all are hybrid functions of multiple functions of subset 1. These functions are more complex than subset one. Therefore, these functions considered to visualize how fast the algorithms reach the optimal solutions. Population size is considered 100 as before, but only for one trial and observe best values over 1000 generations. For all experiments population is initialized with uniformly distributed random values covering solution domain.

Convergence results on hybridized functions f15 to f25 are displayed in Figure 5.3. PSO-CATV shows slow convergence at beginning as it should be, because initially PSO-CATV explores and as gradually decreases cognitive avoidance component it start exploiting. Hence, convergence of PSO-CATV is observable almost after 50% of generations completed. However, exploration for other three algorithms ends early. Though they exploits thereafter, best fitness value does not improves significantly and after a while remains constant. It is also notable that despite being PSO-CATV starts with slow convergence, but most of the cases at the end of execution gradually improves best fitness and results better than all three competitors. Though in functions f15, f20 and f21 shows exception to this observation as PSO-CATV lagging behind some these competitors, exploration of solution space is observable. PSO-CATV shows extensively very high convergence rate in f24 and f25, it leaves all three algorithms far behind both in terms of best result and exploration. Overall performance of PSO-CATV is better than other three algorithms in terms of convergence.

5.4.4 Analysis of Community Detection with PSOCA

The performance of community detection with PSOCA is compared with other three variants of PSO: PSO-TVIW [180], PSO-TVAC [158], OLPSO [221]. Three real-world



CPU Time (in seconds)

Algorithm / Data Sets	PSOCA	PSO-TVAC	PSO-TVIW	OLPSO
Blogger: Data Set 1	2.719	2.959	2.795	52.923
Blogger: Data Set 2	52.466	410.199	349.341	2772.105
Iris: Data Set	2.216	2.622	2.425	19.176

Figure 5.4: Comparison of communities predicted with PSOCA and other variants of PSO on three datasets.

datasets are considered to prepare multiple featured network and given as input to proposed approach for community detection. Two of them i.e. Blogger1 and Blogger2 are related to blogger networks and third one is standard Iris data. Blogger's data set 1 is collected from the UCI Machine Learning Database [60]. The data set consists of a table of 100 bloggers with 6 features. Iris data set [53], which contained 150 instances and 4 features, is also collected from the same source. Blogger's data set 2 is collected from the Arizona State University [3]. The data file has been processed to obtain associated tags. This file is further processed to form a matrix in which the columns are the tags from the data set, and the rows represented the bloggers. Entries of the first 100 bloggers from this dataset have been used for analysis in the experiments. Experiments have been done for 10 trials on each dataset and considered the best result among these trials for comparison.

To mark the best results, three cluster quality measuring metrics namely Dunn index, Davies-Bouldin index and rand index are considered. Dunn index has two different aspects intra i.e. measures how dense the cluster is and inter i.e how sparse clusters are from each other. Davies-Bouldin index also indicates similarity within the cluster and for a better cluster quality, its value has to be minimum. Rand index states accuracy of clusters formed i.e. how appropriately the elements belong to one cluster. Rand index may have value in between 0 and 1. Rand index tending to 1 indicates good and accurate clustering.

The results yielded on each of these datasets with proposed approach are presented in Figure 5.4. In Blogger's data set 1, best results were generated for 2 communities. The first community consisted predominantly entries of bloggers with a higher degree of education, few bloggers with medium level of education, and very few with low level of education. Bloggers in this community predominantly blogged about political topics. Second community predominantly had entries of blogger with medium level of education and few with low level of education and fewer with high level of education. Majority of bloggers in the second cluster blogged about news. The data set was efficiently clustered into 2 sets viz. professional bloggers and seasonal (temporary) bloggers. With this variant of PSO value found for Dunn index (intra) was almost 3 and for rand index was almost 1. The Dunn index (inter) and Davies-Bouldin index of 1.5 was achieved.

Blogger's data set 2 was grouped into 15 communities some of which were singleton sets, these singleton sets were justified because some bloggers in the data set blogged about topics different from all others. An appreciable rand index of 0.709 was achieved in this case. In Iris dataset, rand index is approaching the ideal value of 1. Iris and Blogger's data set 1 have taken lesser computation time. Though both Blogger's data have same number of instances 2nd data set takes much higher time than 1st one. This higher time is because of number of features present in the dataset.

5.5 Conclusion

In this chapter, an algorithm is proposed for detecting communities in multiple featured networks. The algorithm maximizes an objective function that is designed specifically for community detection in multiple featured networks using an improve version of PSO. Proposed a cognitive avoidance mechanism to improve performance of PSO, which is referred as PSOCA. Proposed approach is evaluated with two sets of benchmark functions comprising total of 29 benchmark functions. Furthermore, quality of communities identified on real-world networks using proposed approach is also compared with other variants of PSO. The analysis revealed following interesting facts.

- The novel cognitive avoidance mechanism helps avoiding particle's unfruitful moves. New additional component called cognitive avoidance is introduced to the velocity equation of PSO.

- Cognitive avoidance coefficient controls the effect of this component. The approach tracks personal worst positions and directs particle to avoid movement towards such positions.

- The concept of linearly decreasing cognitive avoidance coefficient is also introduced (PSO-CATV) to negotiate the negative effect and maintain balance between exploration and exploitation.

- Empirical result shows performance of two proposed cognitive avoidance schemes PSOCA and PSO-CATV are better than other variants of PSO and jDE.

- It is worthwhile to mention that additional cognitive avoidance component has definitely improve convergence of PSO as most of the cases PSO-CATV is able to reach more nearer to the optimal value. Moreover, community detection with PSOCA faster than other variants of PSO.

- Quality of communities identified with PSO-TVAC and PSOCA are similar. PSO-TVAC takes more time due to the slow convergence to optimal value. Although OLPSO identifies good communities, but it takes very high computation time.
- In conclusion if good communities as well as computation time is the goal then PSOCA will be better option for detecting communities.