Detection of community in social network using adaptive differential evolution

Reena Rani[1], Jyotsna Saini[2]

[1] *Assistant Professor, computer science and engineering,JMIT, Haryana, India*
[2] *Student, computer science and engineering,JMIT, Haryana, India*

ABSTRACT

Nowadays, ample techniques are innovated to detect communities in social network as several people are interested in online social networks. Online communities are detected through cultural algorithm to solve the problem of optimization using knowledge repository but there are some limitations of this technique which is overcome by other method. This paper deals with the community detection in social network using adaptive differential evolution and compare the result with 3 well- known algorithms in this field and this method dramatically reduce the time of reaching to final answer.

Keyword *social network, adaptive differential evolution, cultural algorithm, community detection, nmi, fitness function, optimization.*

1 INTRIDUCTION

Nowadays, networks like internet, we, social networks, biological networks, mobile and wireless networks become crucial part of human's life. These networks comprise of social agents and one can extract large amount of data sing these networks. It is used in physics, computer science, biology and politics and sociology.

Network is nothing but a graph, G (V, E), where V is set of vertices and E is a set of edges that link the nodes. This graph is formed using adjacency matrix. Let A be an adjacency matrix then A (i, j) is 1 if there is a direct link between two nodes i and j and 0 if there is no edge.

Social network is the combination of connected communities. To understand the structure of the network, discovering the communities is vital. Therefore, various techniques and strategies have been conducted to find such communities efficiently. In short, a community is collection of nodes that has more internal links than external.

Community detection is a kind of NP- HARD problem and only evolutionary algorithm can solve this problem efficiently.

In this paper, we intend to detect community using adaptive differential evolution algorithm. The central idea of this algorithm comes from the fact that it detects the communities quickly, even faster than cultural algorithm, DE-CD and genetics. The central part of this algorithm is belief space which stores the information about sources and destinations. This belief space can be formed using different source of datasets.

The rest of the paper is organized as follows. In section 2, the main literature review is done in the same area. In section 3, merits and demerits of several algorithms are compared and contrast and finally the list of sources.

2. RELATED WORK

    A. *Fitness Function:*

        Let $S = (I, J)$ be sub-matrix of A, where I is a subset of the rows $X = \{I1. . . IN\}$ of A, and J is a subset of the columns $Y = \{J1, . . . , JN\}$ of A.

        Let aiJ denote the *mean value* of the ith row of the S, and aIj the mean of the jth column of S. More formally,

$$aiJ = 1/|J| \_j{\in}J \; aij, \text{ and } aIj = 1/|I| \_i{\in}I \; aij$$

        The *volume vS* of a sub-matrix $S = (I, J)$ is the number of 1 entries aij such that $i \in I$ and $j \in J$, that is $vS = \_i{\in}I, j{\in}J \; aij$.

Given a sub-matrix $S = (I, J)$, the *power mean of S of order r*, denoted as $\mathbf{M}(S)$ is defined as

$$\mathbf{M}(S) =\_i \in I \ (a_{iJ})r//I/$$

A measure based on volume and row/column mean, that allows the detection of maximal and dense sub-matrices, can be defined as follows. Given a sub-matrix $S = (I, J)$, let $\mathbf{M}(S)$ be the power mean of S of order r. The *score* of S is defined as $Q(S) = \mathbf{M}(S) \times vS$. The *community score* of a partitioning $\{S1, \ldots Sk\}$ of A is defined as

$$CS = (k\_i)Q(Si)$$

The problem of community identification can be formulated as the problem of maximize CS. It is worth to note that higher values of the exponent r bias the CS towards matrices containing a low number of zeroes. In fact, it amplifies the weight of the densely interconnected nodes, while reducing those of less connected in the computation of the *community score*. In the experimental result section we show that when the modular structure of the network is not well defined, higher values of r help in detecting communities.

B. The *Normalized Mutual Information*

The Normalized Mutual Information is a similarity measure proved to be reliable by Danon et al. [5]. Given two partitions A and B of a network in communities, let C be the confusion matrix whose element Cij is the number of nodes of community i of the partition A that are also in the community j of the partition B. The normalized mutual information $I(A,B)$ is defined as :

$$I(A,B) =-2\_cA_{i}=1\_cB_{j}=1 \ Cij log(CijN/Ci.C.j)\_cA_{i}=1 \ Ci.log(Ci./N) +\_cB_{j}=1 \ C.j log(C.j/N)$$

Where cA (cB) is the number of groups in the partition A (B), $Ci.$ ($C.j$) is the sum of the elements of C in row i (column j), and N is the number of nodes. If $A = B$, $I(A,B) = 1$. If A and B are completely different,

3. PSEUDOCODE

Input = graph, adjacency matrix
% initializes the network state,

$$NS = \{N_1, N_2, \ldots, N_n\}, n = number\ of\ nodes$$
$$N_i = \{Neighbors\ of\ Node\#i\}, 1 \le i \le n$$

% define an individual $\ I = \{i_1, i_2, \ldots, i_n\}$
% initializes the population,

$$P = \{I_1, I_2, \ldots, I_p\}, \quad p = size\ of\ the\ population$$
$$I_i = Represent(NS), 1 \le i \le p$$

% evaluate fitness, $Fitness(P)$
% sort by the fitness values
% Select population to update the culture

$$SP = \{P_1, \ldots, P_s\}, \quad s = size\ of\ the\ selected\ population$$

% initializes the belief space, $\{BS = (SP)^T$

% set the iteration, $Iteration = 1$

While condition is not met
% generate the population, $For\ i = 1\ to\ p$

- Pick three agents $\mathbf{a}, \mathbf{b}$, and $\mathbf{C}$ from the population at random, they must be distinct from each other as well as from agent $\mathbf{x}$
- Pick a random index $R \in \{1, \ldots, n\}$ (n being the dimensionality of the problem to be optimized).
- Compute the agent's potentially new position $\mathbf{y} = [y_1, \ldots, y_n]$ as follows:
- For each i, pick a uniformly distributed number $r_i \equiv U(0,1)$
- If $r_i <$ CR or $i = R$ then set $y_i = a_i + F \times (b_i - c_i)$ otherwise set $y_i = x_i$
- (In essence, the new position is outcome of binary crossover of agent $\mathbf{X}$ with intermediate agent $\mathbf{z} = \mathbf{a} + F \times (\mathbf{b} - \mathbf{c})$.)
- If $f(\mathbf{y}) < f(\mathbf{x})$ then replace the agent in the population with the improved candidate solution, that is, replace $\mathbf{X}$ with $\mathbf{Y}$ in the population.

% evaluate fitness, $Fitness(P)$

4. RESULT

The simulation studies involve the deterministic communities with 35 nodes as shown in Fig.1. The proposed differential evolution algorithm is implemented with MATLAB. Proposed algorithm is compared between two metrics algorithm one is cultural algorithm and other is cultural DE. Simulation time is calculated through the CPUTIME function of MATLAB

The poly line showed in Fig. 1 is iteration wise optimization in which output based on cultural algorithm is based on red line whereas output based on cultural- DE is represented by blue line. Iteration is represented on x-axis and community score is represented on y- axis. It is clear from the given poly line graph that the out of cultural- DE is far better as compare to simple cultural algorithm.

It clearly shows in Fig. 2 that cultural DE is far better as compared to Cultural algorithm. Community score for cultural algorithm is 30 while for cultural-DE the data accounts for near 35. It is the overall output of the algorithm.
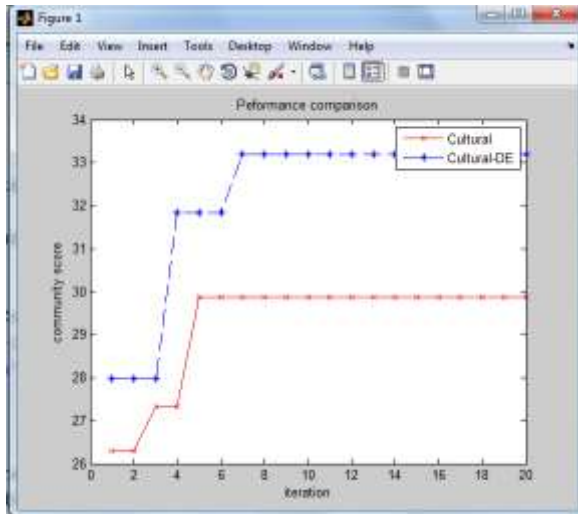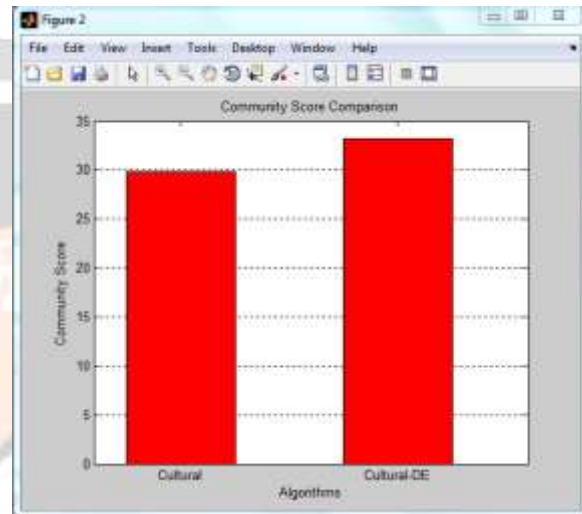

Fig 1 (iteration wise fitness optimization)   **Fig 2 (overall output)**

4. CONCLUSIONS

In conclusion, it is observed that cultural-DE outweighs the performance of cultural algorithm. Its time complexity is far better than others. This algorithm is checked on only Karate dataset and in future its complexity can be checked on different or other datasets in future. On the other hand, we can also use the combination of cultural algorithm with self adaptive differential evolution algorithm to get best result from proposed algorithm.

6. REFERENCES

[1]C. pizzuti, *"GA- A genetic algorithm for community detection in social networks"* parallel probl. Solving from Nat.x, pp. 1081-1090, 2008.

[2]Lu Zongqing *"algorithms and applications for community detection in weighted networks"* 2015.

[3]M. Girvan and M. E. J Newman *"community structure in social and biological network"*, Proc. Natl. Acad. Sci., vol 99, no. 12, pp. 7821- 7826, 2002.

[4]Zadeh P. M. And Ziad Kobti *"community detection in social network using cultural algorithm"* 2015.