Community Detection in complex social networks

by

Khaled Ahmed

Workshop on Intelligent System and Applications (ISA’17)
Overview

- Introduction
  - Motivation
  - Problem Definition
- Related Work
- Proposed Approach (Krill, Chicken, Elephant)
- Results and Discussion
- Conclusion and Future Works
Introduction

The rapid increase on the social networks present an urgent need for identifying the community structure.

A social network is a set of nodes which represent users or profiles and a set of edges, which represents the interaction between nodes; it could be internal or external connectivity.
Can we get Valuable insights for these social networks? One of these Valuable insights is ‘Community detection‘.

Why Community detection?
- It helps in many analysis the social networks:
  - citation network might represent related papers on a single topic, represent pages of related topics, easy to visualize and understand.
Problem Definition

- Finding Communities in complex network such as Online Social network is a difficult task, many challenges.

- **Scalability**: such networks can be huge, often in a scale of millions of actors and hundreds of millions of connections, Existing Community detection techniques might fail when applied directly to networks of this size.

- **Heterogeneity**: In reality, multiple relationships can exist between individuals, and multiple types of entities can also be involved in one network.

- **Evolution**: Social network emphasizes timeliness, which makes the network dynamic and changes over time.

- **Evaluation**: the task of comparison and evaluation different work in community detection is also a challenging process, because of the lack of ground truth for many social computing tasks.
Related work

- There are common algorithms used for community detection as Benchmarks: Discrete BAT Algorithm, Artificial fish swarm algorithm, Infomap, Fast Greedy, label propagation, Multilevel, and Walktrap.
Contributions of our Work

- Change the swarm algorithm domain
- Algorithm redesign to fit for community problem
- Present a new discrete algorithm
- Comparing our results with 7 popular algorithm over four benchmarks.
The proposed three models (Papers)

An Efficient Approach for Community Detection Based on Elephant Swarm Optimization Algorithm

Khaled Ahmed 1,2, About Elham Hassanien 1,3 and Hisham Elsaid 1

1 Faculty of Computers and Information, Cairo University, Giza, Egypt
2) College of Information Science and Engineering, Donghua University, Shanghai, China
3) Scientific Research Group in Egypt (SRGE), Cairo, Egypt

Abstract—Complex social networks analysis is an important research trend, which basically based on community detection. Community detection is the process of dividing the complex social network into a dynamic number of clusters based on its edges connectivity. This paper presents an improved classic Elephant Swarm optimization Algorithm for Community Detection Problem (ESOP) as an optimization approach for solving complex networks community detection problem. ESOP can define dynamically the number of communities within complex social networks. Experiment prove that ESOP can handle the community detection problem and it can handle non-overlapping communities with high accuracy and quality measures of NMI and MODularity over four benchmark datasets and compared with eight common community detection algorithms.

Index Terms—Nature inspired intelligence; Elephant swarm optimization; Community detection; Social networks analysis.

1. Introduction

Social networks is a graph of nodes and edges which represent actors or users in the social community and interaction among them [1]. Social networks analysis is an important research which helps in many medical, economic, educational and social disciplines [2]. Community detection in complex social networks is an urgent analysis method causing the daily rapid increase in social networks and in the number of users.

Community detection is the process of defining a dynamic number of clusters within complex social network and divide the network into a set of clusters based on its edges connectivity, which means edges with high density and high mutual interactions will be in the same cluster. Community detection can help in dividing the network into different parts, each cluster has its own characteristics such as week sport, language and education [3].

Many researchers tackle this challenge of discovering insights from the complex networks using the multilevel techniques and approaches such as spectral clustering, Graph Partitioning, Hierarchical Clustering, Clique-based Graph Cuts and Seed-based algorithms [4].

3.1 The proposed approach

The proposed approach is focused on four popular cited benchmark data sets with diferent sizes of small, medium and large data sets such as Zachary karate clubs, Dolphins Dolphin, American college football and Facebook. Experimental results are measured with quality measurements NMI, Modularity and Ground truth. ESOP’s results are compared with eight well-known community detection algorithms such as Label Propagation, Advance eigenvector, modularity, ground-truth, modularity, and ground truth.

The proposed approach is one of the four popular cited benchmark data sets with diferent sizes of small, medium and large data sets such as Zachary karate clubs, Dolphins Dolphin, American college football and Facebook. Experimental results are measured with quality measurements NMI, Modularity and Ground truth. ESOP’s results are compared with eight well-known community detection algorithms such as Label Propagation, Advance eigenvector, modularity, ground-truth, modularity, and ground truth.

1 Introduction

Social network is a graph of nodes and edges [1] which can lead to discover many valuable insights. These nodes represent online users’ profiles and edges represent the interactions between nodes [2]. Community structural community is to define complex network’s topology in order to divide this complex network into a set of disjoint communities.
Proposed Approach 1: Krill

Input: A Network \( G = (V,E) \)
Output: Dividing and grouping networks nodes into suitable communities.

Initializes the parameters: initial \( I \) with iteration number, \( I_{max} \) with maximum number of iteration, Swarm Size, initializes krill in search space with its values and features: Current state and fitness value, Krill positions, Food position;

while \( i < \text{Krill.Count()} \) do
    store the best solution at iteration \( i \);
    for \( \text{krill}[0] \leftarrow 0 \) to \( \text{krill}[\text{herd.size}] \) do
        \( \text{Krill.Post} \leftarrow \text{motion.induced}() \);
        \( \text{Krill.Post} \leftarrow \text{foraging.Motion(Krill.Post)} \);
        \( \text{Krill.Post} \leftarrow \text{Chaotic.diff(Krill.Post)} \);
        \( \text{ApplyGeneticOperators()} \);
        \( \text{UpdatingtheKrillPosition}() \);
    end
    \( i = i + 1 \); until \( i \) less than Max Iterations;
end

return the best solution

Algorithm 1: AKHSO algorithm steps.
Proposed Approach 2: Chicken

Input: A Network G = (V,E)
Output: Dividing and grouping network’s nodes into suitable communities

Initialize chicken swarm in search space with its values and features;
while i < swarm.Count() do
    store the best solution at iteration i;
    for Chicks[0] ← 0 to chicks[swarm.size] do
        if (i == rooster)
            update rooster position using equation 2;
        elseif (i == hen)
            update hen position using equation 4;
        else
            update rooster position using equation 8;
        Updating the swarm Positions;
    end
    i = i + 1; until i less than Max Iterations;
end
return the best solution

Algorithm 1: ACSO algorithm steps.
Proposed Approach 3: Elephant

Algorithm 1: EESO For Community Detection Algorithm

1. **Input**: A network $G = (N,E)$.
2. **Output**: Dividing and grouping the network's nodes into suitable communities.
3. **procedure**: Initializes the parameters: initial with iteration number, $max$ with maximum number of iteration, $N_e$, Swarm Size, number of clan (d), initializes elephants in search space with its values.
4. **While** $t < Max\_Iteration\_number$ 
5. **store** the best solution at iteration $t$.
6. **For** each $elephant \leftarrow Swarm\_size$
7. Sort Elephants Swarm according to fitness();
8. **If** (elephant.fitness() == best fitness)
9. **Then** update its position using equation 3.
10. **else**
11. **Then** update its position using equation 4,5.
12. **end if**
13. separate elephant with worst fitness () using equation 7;
14. **end for**
15. evaluate the new positions and population.
16. $t \leftarrow t + 1$
17. **Until** $t > Max\_Iteration$
18. return the best solution.
19. **end procedure**

![Elephant swarm optimization process diagram](image-url)
Results and discussion
Results and discussion
community detection for Facebook data set.
Results and discussion

community detection for Bottlenose Dolphin data set.
Results and discussion

community detection for Zachary Karate data set.
Results and discussion

community detection for college Football data set.
Conclusion

- We proposed three discrete swarm algorithms for complex social network’s community detection. The locus-based adjacency scheme is used for two reasons first for encoding and decoding tasks, the second for representing a community structure. We used Modularity as an objective function in the optimization process.

- This research presents a quite promising accuracy and high Modularity results for community structure. It presents good results for small and medium size of data sets, although it presents medium results over big data sets.
Future Works

- Improving performance over big data sets.
- Presents hybrid swarm with other optimization algorithm. Presents a comparative study for more than Swarm optimization algorithm over social networks based on experimental results.
Open Research Problems in social networks

- Build Prediction model (Post, behavior of nodes, ..)
- Spam detection
- Information propagation
- Node influence
- Internet of things
- Sentiment analysis over social networks
- Feature selection over social network data
Thanks and Acknowledgement