Research Proposal

Chance-Constraint Optimization for Community Detection in Complex Networks

By
Zar Bakht Imtiaz
M.S - Computer Science
Reg. #: MCS07173049

Supervisor
Dr Mubashir Ahmad

Department of Computer Science

The University of Lahore,
Sargodha Campus
Recommendation Sheet

Certified that the research proposal of

Ms. Zar Bakht Imtiaz,
MCS07173049,
with the title

“Chance-Constraint Optimization for Community Detection in Complex Networks”

has been reviewed and recommended to be presented to
Research Committee for final review and approval.

Supervisor: Dr. Mubashir Ahmad
Undertaking

I certify that research work titled “Improved Community Detection in Complex Network Using Multi Objective Optimization” is my own work. The work has not, in whole or in part, been presented elsewhere for assessment. Where material has been used from other sources it has been properly acknowledged/referred.

____________________
Zar Bakht Imtiaz
MCS07173049
Acceptance by Research Supervisor

I have read the synopsis and agree to supervise Ms. Zar Bakht Imtiaz for the partial requirement of the degree of MS Computer Science. The proposed synopsis is academically, logistically, administratively and financially feasible. The undersigned is agreed/sure that the student is ready for her synopsis defense.

Name of Supervisor: Dr. Mubashir Ahmad

Signature: __________________________

Date: _________________
Approval by Research Committee

Ms. Zar Bakht Imtiaz have submitted the synopsis well in time and have documented this synopsis according to the provided format. Moreover, the student has incorporated all the comments/observations about Research work/thesis as per Department policy, and is ready to defend for her synopsis defense.

Research Committee Member: ____________________________

Signature: __________________________________________

Date: ____________________
# TABLE OF CONTENTS

1. Abstract ........................................................................................................................................... 1
2. Introduction....................................................................................................................................... 2
3. Literature Review............................................................................................................................. 3
4. Problem Statement ......................................................................................................................... 4
5. Research Objective ......................................................................................................................... 5
6. Research Methodology .................................................................................................................... 6
7. Significance of Research ................................................................................................................ 7
8. Scope and Limitations .................................................................................................................... 8
9. Research Plan .................................................................................................................................. 9
10. References .................................................................................................................................... 10
11. Student Profile ............................................................................................................................ 11
1. Abstract

Discovering community with dense intra links in complex network is one of important issue to be considered. Many algorithms are proposed to deal with this problem. Community detection modeled as optimization problem. Single objective optimization methods are insufficient for complex networks. Using multi objective optimization there are still facing some challenges such as high computation cost and parameter tuning while discovering communities in complex networks. We are going to propose a multi objective community detection technique based on chance constrained optimization which requires less parameter tuning and gives higher accuracy. To show the effectiveness of our scheme, experiments will be performed on complex networks in terms of normalized mutual information and modularity in comparison with recent similar approaches.

2. Introduction

A network is group or system of interconnected people/things. Networks are widely used in different fields such as biology, computer science, physics and math. Different nature of network is existing i.e. biological network, technological network, social network, and political election network. These real world networks are also known as complex networks because of high clustering coefficient. In modern network science, community detection has become a hot topic to research. Communities are formed if nodes sharing some characteristics or having certain attitudes in common. Multiple communities exist within a network. Community detection is used in different fields of life such as Recommender System [1], Link Prediction [2], Early Churn Detection [3], Skill Acquisition in Robots [4] and Text Clustering [5].

Community detection algorithms are depending on topology of the network. No single algorithm performs best on all networks. There are two types of communities found in networks, overlapping and disjoint. In overlapping communities, a node belongs to multiple communities where as in disjoint communities, a node belongs to a single community. For overlapping community detection there are different algorithms introduced such as density based link clustering algorithm [6], Algorithm based on density peaks [7] and local spectral clustering introduced for overlapping communities[8]. The major focus of our research is disjoint community’s detection. Community detection consider as an optimization problem. Optimization problem is solved by two ways. One is exact math and other is meta-heuristics. Exact math increase complexity on complex network so moving towards meta-heuristics to obtain optimal solution for community detection. Meta-heuristics methods for community detection are evolutionary and simulated annealing.

Bilal et al. [9] work on real world networks to detect disjoint communities based on the similarity between nodes using evolutionary algorithm. Ma et al. [10] propose evolutionary nonnegative matrix factorization (ENMF) frameworks. Yuanyuan et al. [11] proposed algorithms that adopt quantum inspired evolutionary algorithm (QIEA) to optimize community detection. In this algorithm there is no need of input number of communities beforehand and also work to optimize generation of population. Evolutionary
algorithm not only work on single objective but also multi objective. Cheng et al.[12] improved individual updating strategy which lead to better quality of community. This proposed algorithm worked on local information named as Local Multi Objective evolutionary algorithm (LMOEA). Zalik.et al. [13] also used evolutionary algorithm to detect communities based on multi objective. Algorithm evaluation performed using problem-specific genetic mutation and initialization.

Yu F. et al. [14] improve quality of discover community structure in complex network. By maximizing modularity algorithm combined simulated annealing selections strategy and a biased search which improve accuracy and convergence speed. Xu et al. [15] introduce new community clustering algorithm by hybrid particle swarm optimization (PSO) and stimulated annealing (SA) named as Spectral Clustering Based on Simulated Annealing and Particle swarm optimization (SCBSP). This algorithm target social networks and preprocessing performed on dataset to get better results. By reducing max no of iteration lead towards more efficient results. Simulated annealing technique also used to detect overlapping communities as Sarswat et al. [16] propose a novel two-step approach (CSO–GA–SA). In this approach simulated annealing is used for local search and detect overlapping communities in social network.

Other than above there also other optimization techniques used to deal problem of community detection like particle swarm optimization technique. Rahimi et al. [17] work on this technique and improve search strategy of PSO. They apply genetic operators i.e. crossovers and mutation to improve the efficiency of search. However, PSO algorithm are easy to fall into local optimum in high-dimensional space and problem of parameter tuning.

In this research, we will propose multi-objective chance constrained optimization method which required no parameter. Additionally, compared to above mentioned algorithms this need less computational memory and easy to implement.

3. Literature Review

In recent years, many community detection algorithms proposed which adopt different strategies. In [18] Loop Edges Delete (LED) Algorithm perform best for detecting overlapping communities. For disjoint communities it sets a similarity threshold. If edges have similarity threshold equivalent or above to set threshold, then edges are not removed from community else its removed from network to formed strong linked community. The key problem for this method is to set an appropriate similarity threshold. To get rid of this threshold selection [19] introduced an algorithm EDCD based on edge-deleting with restrictions. It performed best on disjoint communities and it work like LED but deletion of edges performed a different method. While detecting overlapping communities in social network, due to big search space computation cost of algorithm increase. Community Detection in Social Networks Based on Fire Propagation [20] reduce search space. They introduce Fire Spread algorithm (FSA) to detect overlapping communities in social network.

Some researchers also worked on community detection through label propagation (LPA). In [21], deal with problem that during labeling process, the influence of edge weight and node degree are not considered in large scale network. They propose label
propagation based on node importance and label influence in networks (LPA NI). By avoiding randomness of LPA, this algorithm achieves efficient and stable result as compare to original LPA. In [22], authors dealt with problem of traditional LPA that is when labeling the node during labeling process, algorithm treat the node neighbors equally without considering that multiple neighbors have different effect on node. They proposed algorithm named as Stepping LPA-S. This is designed for undirected and unweighted networks to detect non-overlapping community. The work of LPA-S is extend in [23] by using label propagation and fuzzy C-means known as LPA-FCM. The experiments performed on real world network which shows significance improvement. For a weighted and large scale complex network [24] used label propagation strategy and use objective function weighted compactness function of nodes and communities. Their contribution is improvement in accuracy of network partition.

Heuristic algorithms have shown their effectiveness to solve the optimization problems, like GA [25], [26], BPSO [27],[28], [29] and ACO [30]. A lot of work done in community detection through particle swarm optimization algorithm. Chaitanya et al. [27] deal with problem of dynamic neighborhood topology using PSO. They analyze the node instead of focus on graph structures as other done. This algorithm does not need number of communities beforehand. This approach test on real datasets using three metric evaluations. Zhang et al. [31] present a novel algorithm to deal with problem of LPA instability which down the accuracy of overlapping community. They combine PSO and LPA which show efficiency in detection of overlapping community. Cai et al. [32] combine PSO and modularity function to detect community in small networks. This algorithm Q-PSO get higher values of modularity and normalized mutual information than other algorithms but this increase computational complexity in complex network.

However, all these above-mentioned techniques work on single objective optimization. There also multi objective optimization in which more than one objective required to be optimized. Zou et al. [33] introduced a new objective Negative Ratio Association (NRA) with combination of Ratio cut (RC) named algorithm as community detection problem using inverse modeling multi-objective evolutionary algorithm with decomposition in complex network. They performed extensive experiments on complex network to solve community detection problem. To discover community in dynamic networks Zhou et al. [34] propose a multi objective discrete cuckoo search algorithm. In dynamic networks, the structure of network is not fixed therefore, it must ensure that new node that coming in network does not disturb objective function. This algorithm ensures this by using encoding method and also guarantee the diversity of the population by redefining nest location updating operator. Two fitness function used in this method are snapshot cost and temporal cost. In [30] used a meta-heuristic ant colony optimization algorithm (ACO) for community detection. They discover community structure in complex network using objective function community score and fitness with ACO. This algorithm introduced some new characteristics that lead to detect better quality of community detection. In [35] proposed algorithm (MANIA) for multi objective using node analysis. Firstly, node importance analysis used for single objective but in this paper used for multi objective optimization. They experimentally tested and compared against other novel community detection algorithms.
4. Problem Statement

The scale free nature of Complex Networks makes it difficult to detect internally dense communities. Different mathematical and heuristic techniques are present in literature to address the aforementioned problems, however there are still certain challenges. Former have high computational complexity whereas later are limited by parameter tuning [17].

5. Research Objective

The objectives of our research are:

- Detect strong communities in complex network with dense intra links.
- Computation of multi objective fitness function on complex network.
- Handle the problem of parameter tuning.
- Measure the efficiency of proposed algorithm using measures in quantitative form.

6. Research Methodology

In this research a multi objective community detection using chance constraint optimization approach is proposed. This research will be constructive, applied and quantitative in nature. Complex networks are stored in form of adjacency metrics.

**Representation and generation of population:** The first step is initialization of complex network which is done using locus-based adjacency representation (LAR). It generates random number of populations according to population size. There are several advantages of using this LAR as no need to tell number of communities in advance and took linear time to generate random solution. Population generated random such that for each node i, the value of gene $g_i$ is randomly selected among its neighbor’s node.

**Calculation of Fitness function:** Kernel k-means measure the intra density of identified community whereas Ratio Cut measure inter density. By minimizing these objectives for each solution.

KKM calculated as

$$KMM = 2(n - m) - \sum_{k=0}^{m} \frac{L(V_i, V_j)}{|V_i|}$$

Where $m$ belong to number of community in network and $n$ is number of vertices.

RC computed as follows

$$RC = \sum_{k=0}^{m} \frac{L(V_i, \overline{V_i})}{|V_i|}$$

(2)
Where $|V_i|$ is number of vertices in identified community $i$, $V_i \in P$, $\overline{V_i}=P-V_i$. This will send to input to chance constraint optimization and it will return final solution.

The chance-constrained method is one of the major approaches to solving optimization problems under various uncertainties. It is a formulation of an optimization problem that ensures that the probability of meeting a certain constraint is above a certain level.

7. Significance of Research

In real world, community detection has many applications. Detecting networks of fraud website. These websites selling illicit goods tend to link to each other. If we detect one rogue website, the whole network can get exposed. Therefore, detecting communities of these websites are very beneficial.

Efficiency of predictive models are increased by clustering users. Moreover, unknown feature of entities can be computed with a reasonable degree of confidence using community detection. If a user does not give some information but his/her community has then the missing information can be assigned to user.

8. Scope and Limitations

This research deal with specific multi-objective fitness functions and target only complex networks. Using algorithm, only disjoint communities are detected. Does not deal with dynamic networks and cannot detect overlapping communities in network.
9. Research Plan

![Task Planning Table]

10. References


11. Student Profile

Name: Zar Bakht Imtiaz
Father’s Name: Imtiaz Hussain
Registration #: MCS07173049
Cell No: 0300-7693480
Email: f138183@nu.edu.pk

Educational Qualification: *(Starting from the highest degree)*

<table>
<thead>
<tr>
<th>Degree</th>
<th>Year</th>
<th>Division /%age/CGPA</th>
<th>Board / University</th>
</tr>
</thead>
<tbody>
<tr>
<td>BS(CS)</td>
<td>2017</td>
<td>2.73</td>
<td>National University of Computer &amp; Emerging Sciences.</td>
</tr>
<tr>
<td>F.Sc(Eng.)</td>
<td>2013</td>
<td>A</td>
<td>Sargodha Board</td>
</tr>
<tr>
<td>SSC</td>
<td>2011</td>
<td>A+</td>
<td>Sargodha Board</td>
</tr>
</tbody>
</table>

List of Publications: *(Starting from latest)*

<table>
<thead>
<tr>
<th>S. No</th>
<th>Title of Paper</th>
<th>Journal/Conference</th>
<th>Volume &amp; Issue No</th>
<th>Page No</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

List of Courses Passed in M.S Computer Science: *(Starting from 1st semester)*

<table>
<thead>
<tr>
<th>S. No</th>
<th>Name of Course</th>
<th>Semester</th>
<th>GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Advanced Algorithm Analysis</td>
<td>1st Semester</td>
<td>3.75</td>
</tr>
<tr>
<td>2</td>
<td>Advanced Theory of Computation</td>
<td>1st Semester</td>
<td>3.75</td>
</tr>
<tr>
<td>3</td>
<td>Research Methods</td>
<td>1st Semester</td>
<td>4.0</td>
</tr>
<tr>
<td>4</td>
<td>Advanced Computer Networks</td>
<td>2nd Semester</td>
<td>2.0</td>
</tr>
<tr>
<td>5</td>
<td>Advanced HCI</td>
<td>2nd Semester</td>
<td>3.0</td>
</tr>
<tr>
<td>6</td>
<td>Advanced Operating System</td>
<td>2nd Semester</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>CGPA</td>
<td></td>
<td>3.42</td>
</tr>
</tbody>
</table>

(Signature of the student)