iJRASET
International Journal For Research in
Applied Science and Engineering Technology

Volume: 4 **Issue:** V **Month of publication:** May 2016

DOI:

www.ijraset.com
IC Value: 13.98

Volume 4 Issue V, May 2016
ISSN: 2321-9653

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Identifying the Evolving Communities in Dynamic Networks

P. Divya[1], Dr. P. Rajkumar[2]
[1]P.G Scholar, Info Institute of Engineering, Coimbatore, Tamilnadu, India
[2]Associate Professor, Info Institute of Engineering, Coimbatore, Tamilnadu, India

Abstract: The discovery of evolving communities in dynamic multimodal networks with heterogeneous actors is an important research topic possessing challenging tasks. Evolutionary clustering is a recent framework for clustering dynamic multimodal networks introducing temporal smoothness concept inside the community structure detection. Evolutionary based clustering approaches try to maximize cluster accuracy with respect to incoming data of the current time step, and minimize cluster drift from one time step to the successive one. A temporally regularized framework and its convergence property are used to optimize both these two competing objectives, an input parameter that controls the preference degree of a user towards either the snapshot quality or the temporal quality is needed. In this paper the detection of communities with temporal smoothness is formulated as a multi objective problem and a method based on genetic algorithms is proposed. The main advantage of the algorithm is that it automatically provides a solution representing the best trade-off between the accuracy of the clustering obtained, and the deviation from one time step to the successive. Experiments on synthetic data sets show the very good performance of the method when compared with state-of-the-art approaches.

Keywords- ***Evolutionary clustering, complex networks, dynamic networks, community discovery***

I. INTRODUCTION

In social media, large multidimensional data's exploring and gaining interest in mining the community based on information. In Existing many research focus on community discovering based on keywords and entities through affinity calculation unfortunately, the existing community discovery

Process mining approaches do not take into account the hidden aspects of the intentions behind the data sharing in user activities, recognizing and detecting the hot topics in the network about public opinion on the focus of the community discovery.

Networks are modeled as graphs, where nodes represent individual objects, and edges represent interaction among these objects. Individuals in a network interact with each other and exchange information by forming communities. The detection of community structure, i.e. the organization of nodes into groups having many connections inside the same cluster and relatively sparse connections between vertices of different communities, is a fundamental research topic in the study of complex networks..

In this paper a multi objective approach, named DYNMOGA (DYNamic Multi Objective Genetic Algorithm) to discover communities in dynamic networks by employing genetic algorithms, is proposed. The detection of community structure with temporal smoothness, in fact, can be formulated as a multi objective optimization problem. The first objective is the maximization of the snapshot quality that measures how well the clustering found represents the data at the current time. The second objective is the minimization of the temporal cost that measures the distance between two clustering at consecutive time steps. In order to maximize the snapshot quality to measure the goodness of the division on communities of a network, different quality measures, used in the literature to capture the intuition of network community are considered.

II. MATERIALS AND METHODS

A. Evolutionary Events

In evolving communities there are fundamental events used to characterize the evolution of communities . These events are formulated as set of rules:

1) Birth: The emergence of a step community observed at time t for which there is no corresponding dynamic community in D. A new dynamic community Di is created and added

*2) Death***:** The dissolution of a dynamic community Di occurs when it has not been observed. There has been no corresponding step community for at least d consecutive time steps. Di is subsequently removed from the set D.

3) Merging: A merge occurs if two distinct dynamic communities D observed at time t matching to a single step community C at time t. The pair subsequently share a common timeline.

4) Expansion: The expansion or growth of a dynamic community D occurs when its corresponding step community at time t is significantly larger than the previous front associated with community.

5) Contraction: The contraction or reduction of a dynamic community D occurs when its corresponding step community at time t is significantly smaller than the previous front associated with D.

B. Tracking across time steps

1) Tracking Procedure: Feasibility of performing this correspondence process in an efficient manner for graphs containing a large number of nodes and communities .One approach is to formulate this problem as a weighted bipartite matching task, which involves finding the optimal correspondence between the dynamic community fronts and the step communities.

2) Matching Communities: To perform the actual matching between Ct
And the fronts {F1, . . . , Fk0} , we employ the widely adopted Jaccard coefficient for binary sets

$$\text{sim}\,(C_{ta},\ _{Fi}) = |C_{ta} \cap F_i| \,/\, |C_{ta} \cup F_i|$$

C. *Taxonomy of Community Criteria*

Criteria vary depending on the tasks. Roughly, community detection methods can be divided into 4 categories (not exclusive): *1) 1)*

1) Node-Centric Community: Each node in a group satisfies certain properties

2) Group-Centric Community: Consider the connections within a group as a whole. The group has to satisfy certain properties without zooming into node-level

The group-centric criterion requires the whole group to satisfy a certain condition

E.g., the group density >= a given threshold

3) Network-Centric Community: Partition the whole network into several disjoint sets to structure it as network centric items and dimensions. Nodes satisfy different properties like complete mutuality, Reachability of the member, node degree, and relative frequency.

4) Complete mutuality based structure: Clique: a maximum complete sub graph in which all nodes are adjacent to each other

5) Hierarchy-Centric Community: A hierarchical structure of communities will help to retrieve the data and establish a community rapidly.

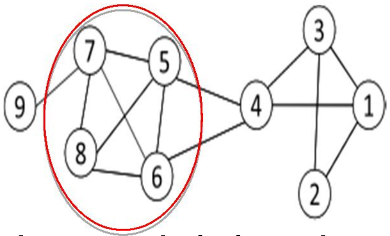
a) Build a hierarchical structure of communities based on network topology

b) Allow the analysis of a network at different resolutions

c) Representative approaches:

i) Divisive Hierarchical Clustering (top-down)

ii) Agglomerative Hierarchical clustering (bottom-up)



Nodes 5, 6, 7 and 8 for form a clique
Fig 2.1 Community detection

III. LITERATURE REVIEW

Analyzing networks and their evolution is recently receiving increasing interest from researchers. In fact, the representation of many complex systems through a static graph, even when the temporal dimension describing the varying interconnections among nodes is available, does not allow to study network dynamics and the changes it occurs over time. Kumar at al [2] studied the evolution of network properties of two large blog sphere networks, and tried to classify members of each network into groups and their changes.

M. Kim And J. Han [1] discovered a particle-and-density based evolutionary clustering method that efficiently discovers a variable number of communities of arbitrary forming and dissolving. We first model a dynamic network as a collection of lots of particles called nano communities and a community as a densely connected subset of particles, called a quasi l-clique-by-clique. Each particle contains a small amount of information about the evolution of data or patterns.

R. Kumar, J. Novak, and A. Tomkins [2] referring to group of vertices within which their connecting links are dense but between which they are sparse. A network community mining problem is concerned with the problem of finding all such communities from a given network. A wide variety of applications can be formulated as NCMPs, ranging from social or biological network analysis to web mining and searching.

Y Chi, X. Song, D. Zhou, K. Hino, and B. Tseng [3] proposed an objective functions for graph clustering called the Q function which allows automatic selection of the number of clusters. Empirically, higher value of the Q unction have been shown to correlate well with good graph clustering's. In this work we show how optimizing the Q function can be formulated as a spectral relaxation problem and propose two new spectral clustering algorithms that seek to maximize.

Y.R. Lin [4] presents an innovative algorithm that deviates from the traditional two-step approach to analyze community evolutions. In the traditional approach, communities are first detected for each time slice and then compared to determine correspondences. Proposing a Facet Net for analyzing communities and their evolutions through a robust unified process.

IV. PROBLEM ANALYSIS

One fundamental issue in social network analysis is the detection of user communities. A community is typically thought of as a group of users with more and/or better interactions amongst its members than between its members and the remainder of the network. In order to capitalize on the huge number of potential users, quality community detection and profiling approaches are needed. It has been well understood that people in a real social network are naturally characterized by multiple community memberships. Thus, it is more reasonable to cluster users into overlapping communities rather than disjoint ones.

V. PROPOSED SYSTEM

Community discovery scheme for a multidimensional network is a proposed model for community discovery in dense multidimensional data. It identify a seed based community structure in a multidimensional network such that the involved items of the entities inside the community interact significantly and mean while they are not strongly influenced by the items outside the community. In this proposal, a community is constructed starting with a seed consisting of one or more items of the entities believed to be participating in a viable community.

A. Community Discovery Framework

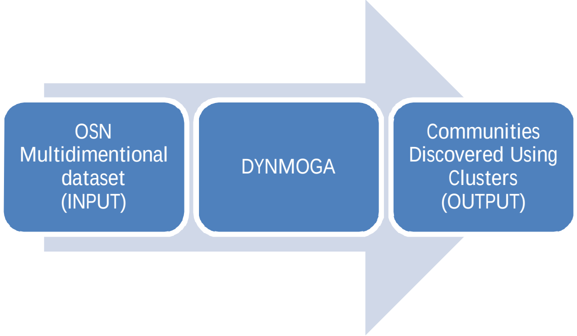


Fig 5.1 Architecture diagram of the community discovery framework

1) *Partitioning Of Network:* In this module, Social data network make use of the eigenvectors of matrix representation of the

network. We show that with certain choices o free parameters appearing in these spectral algorithms that at least within the spectral approximations used here, there is no difference between the modularity and inference based community detection methods or between either and graph partitioning.

2) *Evolutionary Clustering:* Concept of temporal smoothness inside the community structure detection trying to maximize cluster accuracy with respect to incoming data of the current time step and minimize clustering drift from one time step to the successive one.

3) *Modeling Dynmoga:* The detection of community structure with temporal smoothness can be formulated as a multi objective optimization problem. In order to maximize the snapshot quality to measure the goodness of the division in communities of a network, different quality measures are used exploiting the benefits of two functions and exploring the search space, without the need to know in advance the exact number of groups.

B. *Scalability analysis of proposed system*

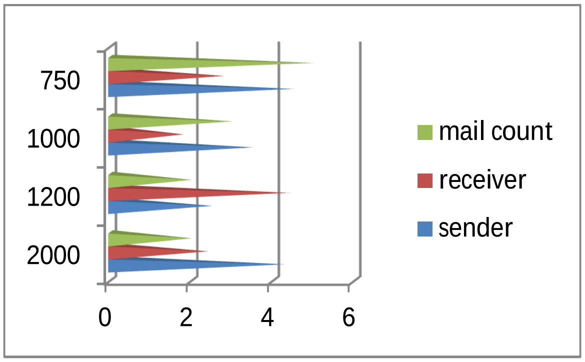


Fig 5.2 Analysis graph for e-mail users

C. *Genetic representation*

Clustering algorithm uses the locus-based adjacency representation proposed. Genes represent nodes of the graph *G = (V, E)* modeling a network *N*, and a value *j* assigned to the *i*-th gene is interpreted as a link between the nodes *i* and *j* of *V*. This means that in the clustering solution found, *i* and *j* will be in the same cluster. A decoding step, however, is necessary to identify all the components of the corresponding graph. The nodes participating in the same component are assigned to one cluster. A main advantage of this representation is that the number *k* of clusters is automatically determined by the number of components contained in an individual and determined by the decoding step

1) Initialization: A population of random individuals is generated such that for each node *i*, the value of *gi* is randomly chosen among one of its neighboring nodes *j*. This means that the edge (*i*, *j*) exists.

2) Uniform Crossover: Given two parents, a random binary mask is created. Uniform crossover then selects the genes where the mask is 0 from the first parent, and the genes where the mask is 1 from the second parent, and combines the genes to form the child. The child at each position *i* contains a value *j* coming from one of the two parents. Thus uniform crossover maintains node connections in the child individual since the edge *(i, j)* exists.

3) Mutation: The mutation operator, similarly to initialization, for each node *i* randomly changes the value of *gi* with one of the neighbors of *i*. This mutation guarantees the generation of a mutated child in which each node is linked only with one of its neighbors.

VI. RESULT

In this section we study the effectiveness of our approach by first employing modularity as objective function that optimizes the snapshot cost, and compare the results obtained by *DYNMOGA* w.r.t. the algorithm

A. *Data Set Description*

Synthetic networks for which the partitioning in communities is known is established with synthetic data sets which is generated with authors .The network is consist of 128 Nodes divided Into four communities of 32 nodes per each community .

B. *Pareto Front Solution Selection*

A characteristic of multi objective optimization is the generation of a set of solutions. Thus, a single solution out of this set must be selected. There has been a lot of research in this decision making problem, and many different approaches have been proposed. As already stated, *DYNMOGA* prefers, among the Pareto front solutions, that having community structure with the higher modularity value, since this concept has been recognized as the most suitable to interpret the intuitive idea of community

C. *System Analysis*

One of the main criticisms in using genetic algorithms, compared with traditional optimization algorithms, is the high execution time required to generate a solution. A major limitation of evolutionary algorithms is the repeated fitness function evaluation that, for complex problems, could often be prohibitive. The problem is exacerbated when large populations of individuals are used, in particular for the multi objective approach. In our method fitness evaluation is rather efficient, thus the main problem comes from the network size.

VII. SYSTEM IMPLEMENTATION

A. Performance analysis

One of the main criticisms in using genetic algorithms, compared with traditional optimization algorithms, is the high execution time required to generate a solution. A major limitation of evolutionary algorithms is the repeated fitness function evaluation that, for complex problems, could often be prohibitive. The problem is exacerbated when large populations of individuals are used, in particular for the multi objective approach. In our method fitness evaluation is rather efficient, thus the main problem comes from the network size.

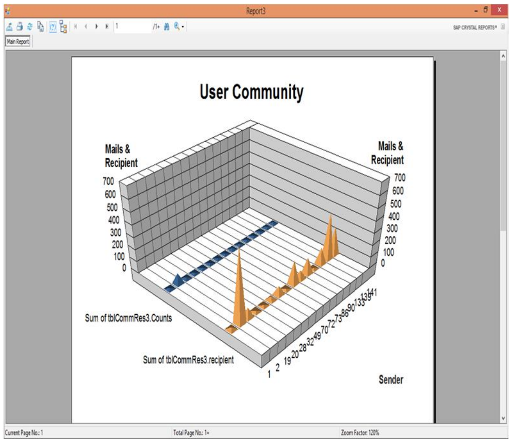


Fig 7.1 Performance Analysis of user community generation against the data available

B. Scalability analysis

Evolutionary-based clustering approaches try to maximize cluster accuracy with respect to incoming data of the current time step, and minimize clustering drift from one time step to the successive one. In order to optimize both these two competing objectives, an input parameter that controls the preference degree of a user towards either the snapshot quality or the temporal quality is needed. The detection of communities with temporal smoothness is formulated as a multiobjective problem and a method based on genetic algorithms is proposed

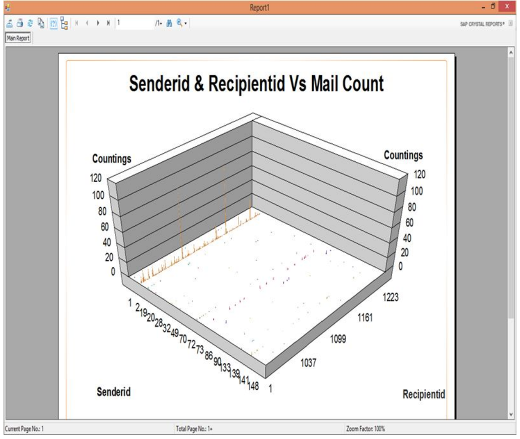


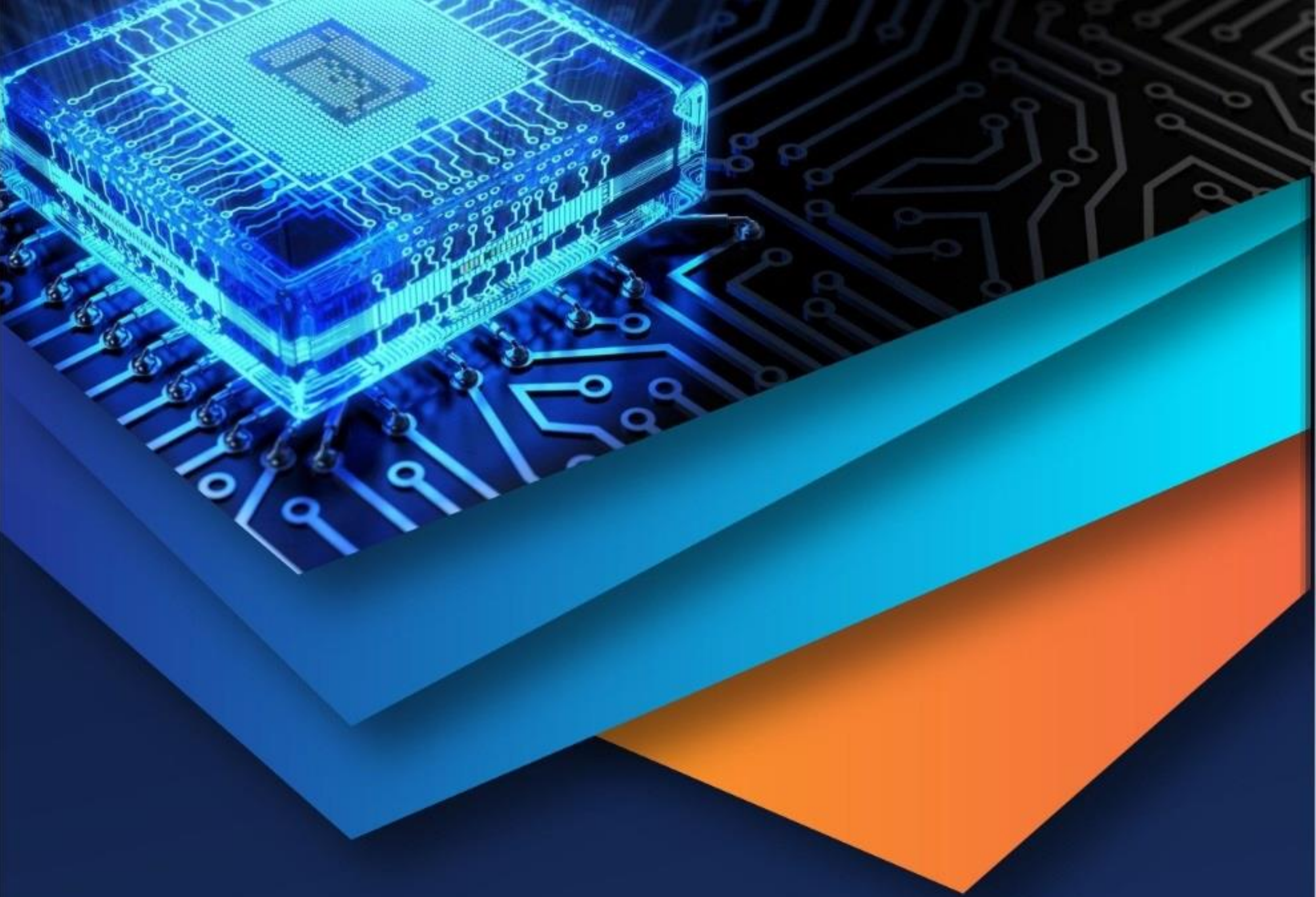
Fig 7.2 Scalability analysis of the proposed system

VIII. CONCLUSION

A multi objective method based on genetic algorithms for detecting communities in dynamic networks has been presented. The algorithm, at each time step, provides the solution representing the best trade-off between the accuracy of the clustering obtained with respect to the data at the current time step, and the drift from one time step to the successive. The proposed approach can be considered as a general framework for evolutionary clustering since changing one of the two objective functions. Thus, if the computing resources are limited, the choice of ***p*** and ***g*** should be done by considering the trade-off between and desired accuracy.

REFERENCES

[1] K. Deb, Multi-Objective Optimization Using Evolutionary Algorithms, Chichester, U.K.Wiley, 2001.
[2] D. Chakrabarthi, R. Kumar, and A. Tomkins, “Evolutionary clustering” in Proc. 12th ACM Int. Conf. KDD, NewYork, 2006, pp. 554-560.
[3] T. Xu, Z. Zhang, P. S. Yu, and B. Long, ”Dirichlet process based evolutionary clustering”, in Proc. 8th IEEE Int. Conf. Data mining, Pisa, Italy, 2008, pp. 648-657.
[4] M. Ehrgott, “Multicriteria Optimization”, 2nd edition Berlin, Germany, 2005.
[5] C. A. C. Coello, G. B. Lamont, and D. A. V. Veldhuizen, “Evolutionary Algorithm for Solving Multi Objective Problems”, Springer, 2007.
[6] . H. Holland, Adaptation in “Natural and Artificial Systems”, Ann Harbor, MI, USA: University of Michigan Press, 2011.
[7] Y. Park and M. Song, “A genetic algorithm for clustering problems”, in Proc. 3rd Annu. Conf. Genetic Algorithms,2012.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)