

Efficient Discovery of Community and Enhanced Privacy Technique in Multidimensional Network

G.Soundarya

MTech Student ,

Department Of Computer Science & Engineering
Madanapalle Institute of Technology & Science
Madanapalle, India

M.Kirankumar

Assistant Professor,

Department Of Computer Science & Engineering
Madanapalle Institute of Technology & Science
Madanapalle, India

Abstract— the goal of proposed system is to develop a group having a particular characteristic in common underhand plan scheme in a complicated or intricate network for data mining. In social media networks data made up of multiple dimensions items like as stories, tags, images, quotes and users. The proposed system showing curiosity in finding action of finding some entities or something. Performing an activity together is significantly great on these media networks. Which two documents are cited together by other documents in network? major involvement of this system is to propose a basic structure underlying a system to find associate something closely with group community in a multidimensional network by making of a judgment between two items in the same or diverse types of entities from the network. To determine the most probable event on each item in dimensional network and verify their values to produce groups from a set of seed entities and also we can evaluate quality of generated communities by using proposed algorithm.

Keywords-community;multi-dimensional network; tensors; affinity calculation

I. INTRODUCTION

In recent times, there are rising interests in studying and analyzing large networks, such as social networks, genetic networks and co-citation networks. In these networks, every node is an item equivalent to a dimension or an entity in a network. every edge indicates a connection between two nodes, for example, a contact between two users in a social network, an communication between two genes in a genetic network, and a reference between two papers or two authors in a co-citation network. Analyzing these networks permit us to understand their topological properties and structural organization. One of such objectives is to identify communities or modules in large networks. One approach is to partition the Network into sub-networks so that nodes in each sub-network are closely associated while nodes in diverse sub networks are loosely connected. For example, the concepts of edge between-ness and modularity, and the Maximum-flow-minimum-cut are used to partition the network. Spectral methods are anticipated to partition the whole network to discover communities. In some applications, the main concern is determine a specific community containing a set of of concern rather than to partition the entire network several groups. For example, we are interested in a

particular sharing group in a social network, or a set of genes along a pathway in a genetic network. To find community structure in a network, Bagrow and Bolt developed an algorithm to search the structure of a starting with a seed node. Clauset proposed a local modularity measure and then developed a seed-based community discovery algorithm. In Luo et al. designed new modularity measure and developed a KL-like algorithm to explore such community structure. these methods were developed for uni-dimensional networks. In this project, we are interested in such a seed-based community discovery scheme for a multidimensional network. There are many data mining information retrieval applications in which there are dimensions/entities in networks.

II. MATERIALS AND METHODS

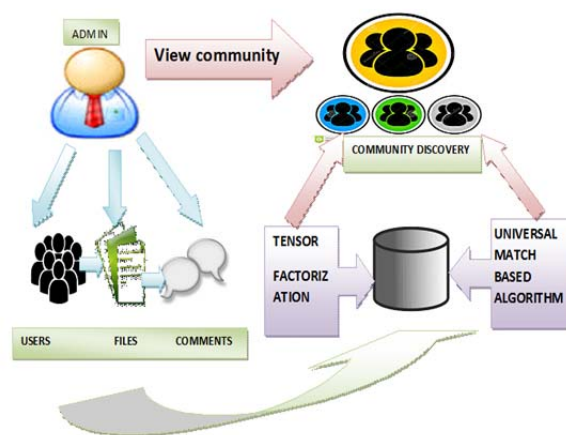


Figure1: Community Discovery in Different Dimensions

Here we are using some modules for implementation

Community Discovery Module
Tensor factorization Module
Affinity Calculation Module
Probability distribution Module

A. Community Discovery Module:

In this module we to find communities based on minmax cut principle, i.e., minimize the connections

between communities while maximize the connections within a community. The basic idea is reordering the adjacency matrix to find dense diagonal blocks, each of which represents a dense community.

B. Tensor factorization Module:

In this module, the interactions between different entities are represented as a set of tensors, some of which have overlapping dimensions. Then it decomposes these tensors into matrices simultaneously using KL-divergence as a measure of approximation cost. Since the KL-divergence is used, the factorization matrices can represent the prior probabilities of different communities and the conditional probabilities of each item in these communities. Based on them, the posterior probability that each item belongs to a particular community can be computed and then the final community results can be obtained.

C. Affinity Calculation Module:

In This Module we calculate the affinity based on calculation of probabilities of visiting other items in a network from a given set of items. Motivated by the idea of topic-sensitive PageRank and random walk with restart, we consider a random walker chooses randomly among the available interactions among the items in different dimensions, and makes a choice with probability a going back to a set of items in the current community.

D. Probability distribution Module

In this Module, we obtain Probability distribution visit each item in each dimension in the multi-dimensional network. These probability distributions can be viewed as an affinity vector because it indicates the affinity of the items in each dimension to the items in the current community. Based on their probability values, we can determine the candidate items in different dimensions that are closely related to the current items in the community.

III. UNIVERSAL MATCH BASED ALGORITHM

A. Algorithm:

Step1: The algorithm begins with grouping arrangement, amid which all nodes that have not yet been assembled are thought seriously about, in Step2: In like fashion, two nodes (v1, v2) with the greatest similitude of their neighborhood names (LSv1,LSv2) are assembled together (3).

Step3: Their neighbor marks are altered to be the same instantly so that nodes in one gathering dependably have the same neighbor names.

Step4: At that point nodes having the most extreme closeness with any node in the grouping are bunched into the grouping till the group has ` nodes with distinctive delicate names (2).

Step5: From that point, the calculation continues to make the following gathering. On the off chance that less than ` groups are left after the last bunch's arrangement, these leftover portion groups are grouped into existing gatherings as indicated by the similitude's in the middle of groups and gatherings

Step6: In the wake of having shaped these groups, we have to guarantee that every group's individuals are unclear as far as neighborhood data.

Step7: Consequently, neighborhood marks are altered after each gathering operation, so that names of nodes can be likewise redesigned instantly for the following gathering operation.

Step8: This adjustment procedure guarantees that all groups in a gathering have the same neighborhood data.

B. Equations:

- 1) Estimating conditional probabilities:

$$P(s, t) (i_{j_1(s)}, i_{j_2(s)}, \dots, i_{j_s(s)}) = \text{Prob} \left[X_k^{(j_t(s))} = i_{j_t(s)} \mid \underbrace{X_{k-1}^{(j_u(s))} = i_{j_u(s)}}_{\forall u=1, \dots, l_s, u \neq t} \right],$$

- 2) calculate the affinity between two items

$$x_v = (1 - \alpha) \left[\frac{1}{|N_v|} \sum_{s=1, j_t(s)=v}^S \mathcal{P}^{(s,t)} \otimes_{u=1, u \neq t}^{l_s} x_{j_u(s)} \right] + \alpha z_v,$$

- 3) Calculate neighborhood label similarity.

$$NLS(v_1, v_2) = \frac{|LS_{v_1} \cap LS_{v_2}|}{|LS_{v_1} \cup LS_{v_2}|}$$

IV. EXPERIMENTAL RESULT AND ANALYSIS

In this area, we introduce some exploratory assessment of the proposed calculation, which uncover an awesome change in the universal match based algorithm .here we are observing changes against the number of items in different communities like one community, two community , three community etc. we are observing each and every item in different dimensions and finally we are overlapping/discovering the communities having similar/same properties.

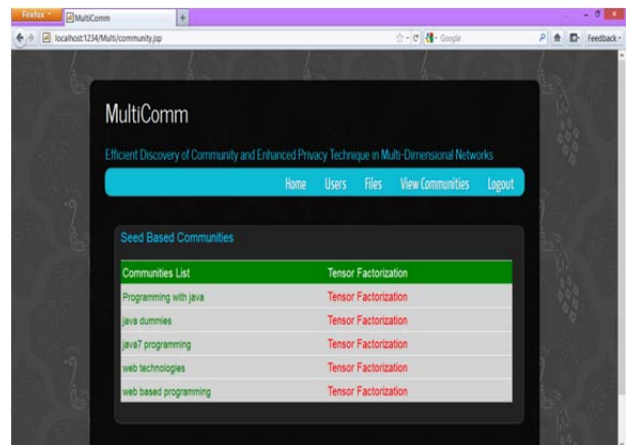


Figure 2: snapshot for community discovery

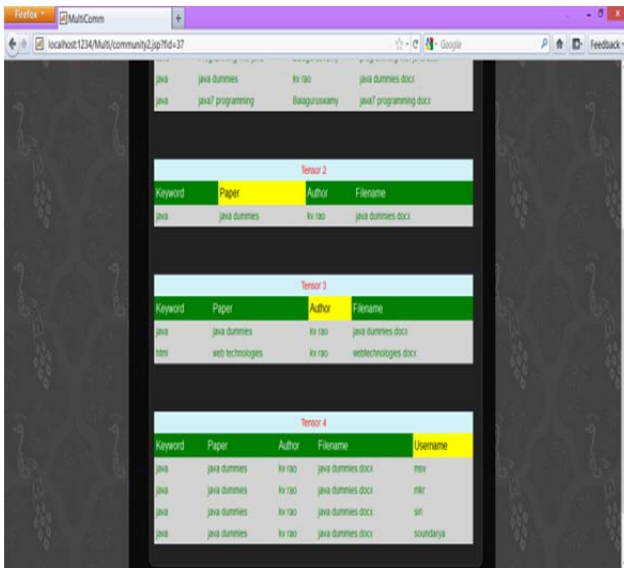


Figure 3: snapshot for individual community

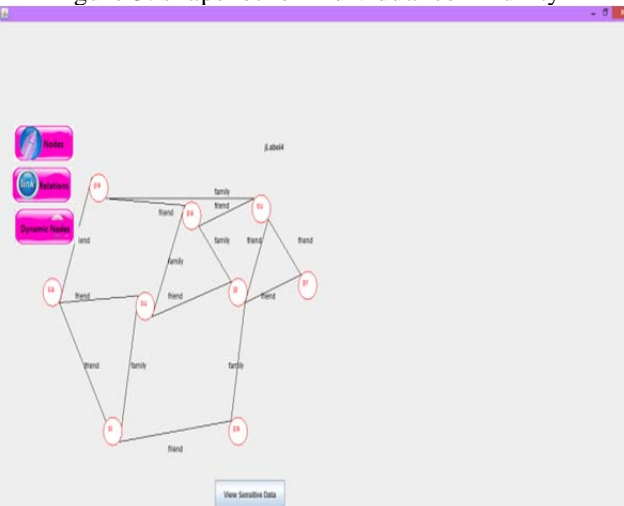


Figure 4: snapshot for Graphical results for different nodes

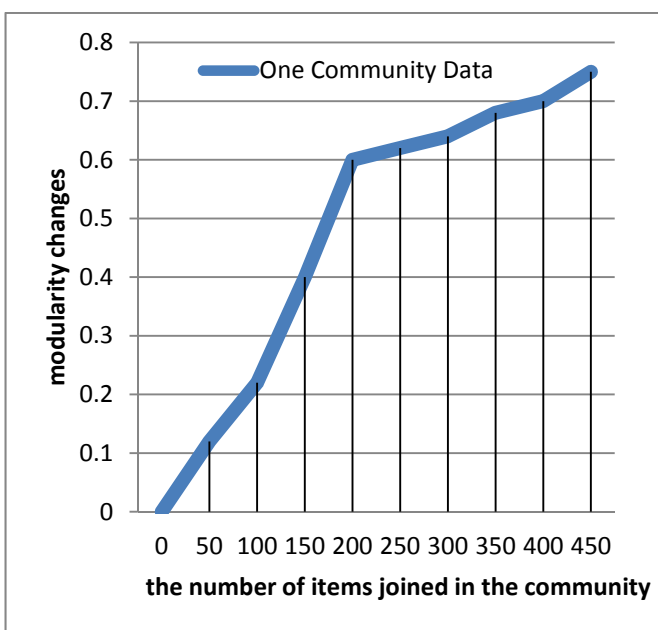


Figure 5(a)

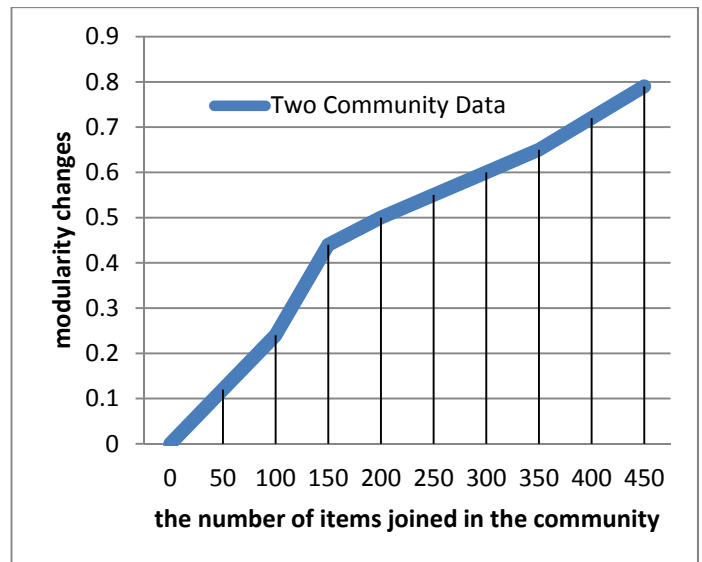


Figure 5(b)

Figure 5: the local modularity changes against the number of items joined in the community
 (a) One community data generated
 (b) Two community data generated

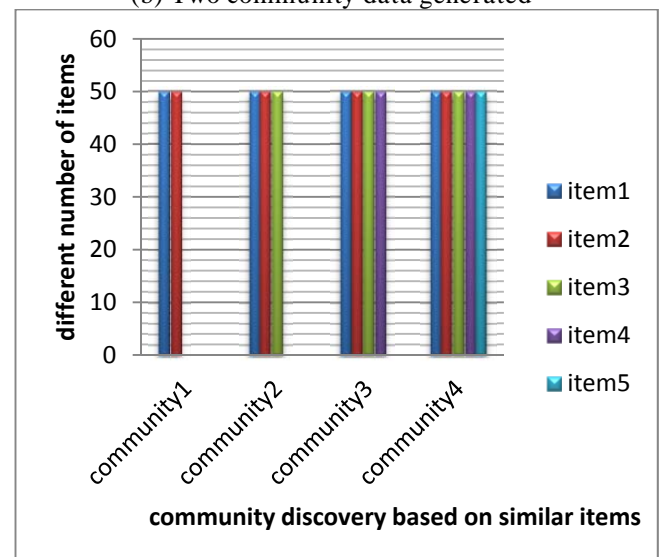


Figure 6: Community Discovery Based on Number of Items

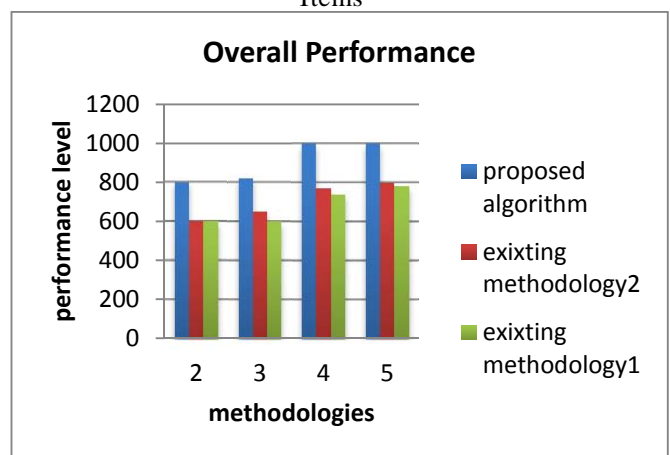


Figure 7: Overall Performance

V. CONCLUSION

In this paper, we have proposed a structure to focus groups in a multi-dimensional system in view of likelihood dissemination of every measurement/substance figured from the system Both hypothetical and exploratory results have shown that the proposed calculation is proficient and powerful. which might be applicable for the opinion analysis and decision making in large business and marketing services.

VI. FUTURE WORK

In future developments there is the plan to systematically investigate to what extent our approach reveals useful for application to real world and computer generated networks .In particular, the aim will be about large scale networks ,for which it may also be unknown the 'true partitioning. In this case the only possible check would be the comparison with results, as provided by different fast algorithms.

ACKNOWLEDGMENT

This is the project done by me in our college as regular curriculum of MTech degree. This project was funded by MITS College of Engineering.

REFERENCES

- [1] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks". *Proc. Natl. Acad. Sci.USA* 99, 7821–7826 (2002).
- [2] G.Palla, I.Derenyi, I.Farkas, T.Vicsek. "Uncovering the overlapping community structures of complex networks in nature and society". *Nature*, 2005, 435(7043): 814-818.
- [3] M.E.J.Newman. "Modularity and community structure in networks". *Proc Natl Acad Sci USA*, 2006, 103(23):8577-8582.
- [4] F.Wu, B.A.Huberman. "Finding Communities in Linear Time: A Physics Approach". *European Physical J. B*, 2004, 38(2):331-338.
- [5] M. Newman and M. Girvan. "Finding and evaluating community structure in networks". *Physical Review E*, 69, 026113 (2004).