Review on Dynamic Community Structures In Multidimensional Social Networks

T.Manjubashini[1], G.S.Geethamani[2]
[1] Dept of Information Technology
[2]Assistant Professor, Dept of Information Technology
[1, 2] Hindusthan College of Arts and Science,Coimbatore, India

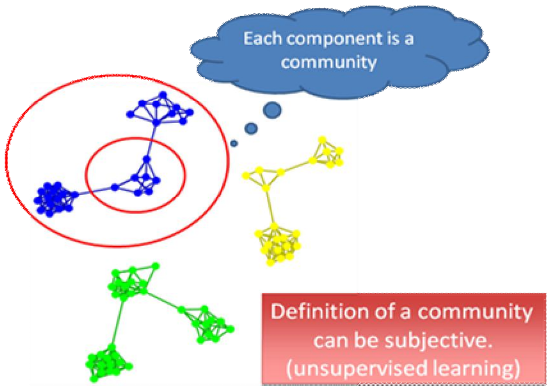
Abstract- *Online Social network is growing to large extent to share information between the different diversity people around the world. The main objective of the proposed system to identify the community in the multidimensional data such as users , Tags , stories , locations ,employment details ,photos and comments . We propose a data mining technique to detect the frequently interacting users based on the common subjects and grouping them in single community. The main incorporation of the work is to identify a seed-based community in a multi-dimensional network by evaluating the affinity between two items in the same type of entity (same dimension) or different types of entities (different dimensions) from the network.*

Social Network Analysis (SNA) is a commonly used method to study social interactions of online groups at an individual level as well as group level. SNA seeks to represent datasets in a form of social networks. In a social network, there are nodes which represent group members, and edges (often referred to as ties) that connect people by means of various types of relations. The strength of the relations is usually conveyed via a weight assigned to each tie. A network representation of social interactions provides researchers with an effective mechanism for studying collaborative processes in online communities, such as shared knowledge construction, information sharing and exchange, influence, support. Because the case examined in this dissertation is online learning communities, the three examples below demonstrate how SNA can be used to study social interactions in online classes.

Keywords- Single Processing, Multiprocessing, Cloud Computing, NeplesAlgorithm.

I. INTRODUCTION

Social networking sites are experiencing tremendous adoption and growth. The Internet and online social networks, in particular, are a part of most people’s lives. EMarketer.com reports that in 2011, nearly 150 million US Internet users will interface with at least one social networking site per month. EMarketer.com also reports that in 2011, 90 percent of Internet users ages 18-24 and 82 percent of Internet users ages 25-34 will interact with at least one social networking site per month. This trend is increasing for all age groups. As the young population ages, they will continue to leverage social media in their daily lives. In addition, new generations will come to adopt the Internet and online social networks. These technologies have become and will continue to be a vital component of our social fabric, which we depend on to communicate, interact, and socialize. Not only are there a tremendous amount of users online, there is also a tremendous amount of user profile data and content online. For example, on Facebook, there are over 30 billion pieces of content shared each month. New content is being added every day; an average Facebook user generates over 90 pieces of content each month. This large amount of content coupled with the significant number of users online makes maintaining appropriate levels of privacy very challenging. There have been numerous studies concerning privacy in the online world A number of conclusions can be drawn from these studies. First, there are varyinglevels of privacy controls, depending on the online site. For example, some sites make available user profile data to the Internet with no ability to restrict access. While other sites limit user profile viewing to just trusted friends. Other studies introduce the notion of the privacy paradox, the relationship between individual privacy intentions to disclose their personal information and their actual behaviour . Individuals voice concerns over the lack of adequate controls around their privacy information while freely providing their personal data. Other research concludes that individuals lack appropriate information to make informed privacy decisions . Moreover, when there is adequate information, short-term benefits are often opted over long-term privacy. However, contrary to common belief, people are concerned about privacy . But managing ones privacy can be challenging. This can be attributed to many things,



II. 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.

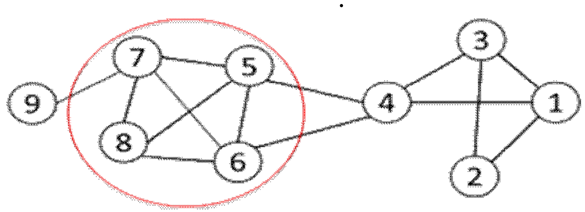
Complete mutuality based structure

Clique: a maximum complete sub graph in which all nodes are adjoining to each other

Location Based Social Networks

A LBSN does not only mean adding a location to an existing social network so that people in the social structure can share location-embedded information, but also consists of the new social structure made up of individuals connected by the link derived from their locations in the physical world as well as their location-tagged media content, such as photos, video, and texts. Here, the physical location consists of the instant location of an individual at a given timestamp and the location history that an individual has accumulated in a certain period. Further, the connection includes not only the physical location or share similar location histories of the person but also the knowledge, e.g., common interests, behaviour, and activities, inferred from an individual's location (history) and location-tagged data . In LBSNs, users can explore places, write reviews, upload photos, and share locations,experiences with others

.



Nodes 5, 6, 7 and 8 form a clique

III. LITERATURE SURVEY

In this literature, community definition is based on the observation that a typical member in a community is linked to many other members, but not necessarily to all other nodes in the community.

A Local Method for Detecting Communities

In this literature, we explore method of community detection that is computationally inexpensive and possesses physical significance to a member of a social network. This method is unlike many divisive and agglomerative techniques and is local in the sense that a community can be detected within a network without requiring knowledge of the entire network.

On the Spectral Characterization and Scalable Mining of Network Communities

In this literature, Network communities refer to groups of vertices within which their connecting links are dense but between which they are sparse. A network community mining problem (or NCMP for short) 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 and/or biological network analysis to web mining and searching. Network communities and their properties based on the dynamics of a stochastic model. Relationship between the hierarchical community structure of a network and the local mixing properties of such a stochastic model has been established with the large-deviation theory. Topological information regarding to the community structures hidden in networks

IV. SYSTEM DESIGN

4.1 Existing System

Community Discovery Scheme is existing works as follows

- Minmax cut principle, i.e., minimize the connections between communities while maximize the connections within a community. Following this principle, it has been shown that the corresponding optimization problem can be relaxed and solved by finding the second lowest eigenvector of its Laplacian matrix.
- Clauset proposed a local modularity measuring the sharpness of a subgraph boundary, and then developed a greedily-growing algorithm based on this modularity for exploring community structure.

Disadvantage of the Existing System

- Systems fails incorporate the information from the Dense Community.
- Information Retrieval for grouping can be expanded to single or two dimensional Networks but for multidimensional performance fails in terms of retrieval speed and data relevancy rate
- Structural Information are often too sparse and weak
- It is difficult to detect the overlapping communities.
- Establish clusters based on edges (user preference) instead of the User.
- It leads to co clustering issues

4.2 Proposed System

Seed-based community discovery scheme for a multidimensional network is proposed model for Community discovery in dense multidimensional Data . It identify a seed-based community structure in a multi-dimensional network such that the involved items of the entities inside the community interact significantly, and meanwhile, they are not strongly influenced by the items outside the community. In our 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. Given the seed item, we iteratively adjoin new items by evaluating the affinity between the items to build a community in the network. As there are multiple interactions among the items from different dimensions/entities in a multidimensional network, the main challenge is how to evaluate the affinity between the two items in the same type of entity (from the same dimension/entity) or in different types of entities (from different dimensions/entities). On the other hand, we need a criterion in order to evaluate a high quality of generated communities by the proposed algorithm, and thus we study a local modularity measure of a community in a multi-dimensional network. data suggest that the proposed framework is able to find a community effectively.

Advantage of the Proposed System

- Proposed algorithm is better in accuracy than the other testing algorithms in finding communities.
- Inter behaviour and intra behaviours of user is obtained
- Multimodal perspective is used to avoid the co clustering issues
- Data Redundancy is eliminated among the communities.

Algorithm to Discovery the Community based Multi dimensional data

Input: Data Source – User Details and Activity formed in terms of Multidimensional data
Process:

Classify the user details and activity based on the different constraints
Constraints has modelled as learning algorithm
Classify the Training data into class based on the attributes of the Dataset
Classify the attribute based on Domain Knowledge and Value types

V. RESULTS AND DISCUSSION

We have proposed a framework (MultiComm) to determine communities in a multi-dimensional network based on probability distribution of each dimension/entity computed from the network. Both theoretical and experimental results have demonstrated that the proposed algorithm is efficient and effective. Performance of the proposed System is determined through the following parameters. Precision and recall are calculated in terms of the ground-truth community. We construct one "ground-truth" community and add noisy interactions in a tensor, and then check how different algorithms can recover this community. There are two parameters to control the data generation. The parameter b is used to control how strong the interactions among items in the community. We also construct two "ground-truth" communities and add noisy interactions in the generated networks. The two communities can be overlapped, i.e., an item can belong two communities

Construction of the Online Social Networking

We have proposed a framework Community overlapping protocol to determine communities in a multi-dimensional network based on probability distribution of each dimension/entity computed from the network.

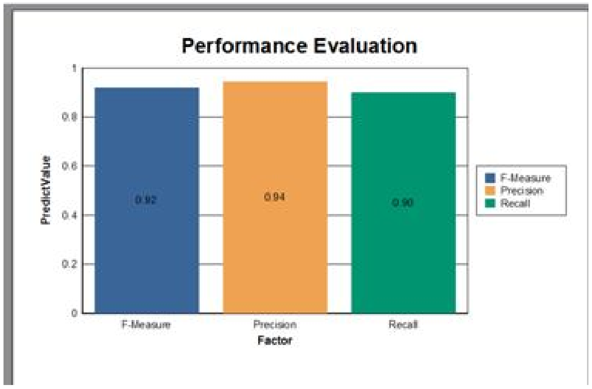


Figure 5.1: Performance Evaluation of the Implemented Community Prediction Algorithms

Experimental results showed in figure 5.1 the proposed framework was able to discover high quality overlapping communities from different perspectives and at multiple granularities, which can be used to facilitate different applications, such as group advertising and marketing.

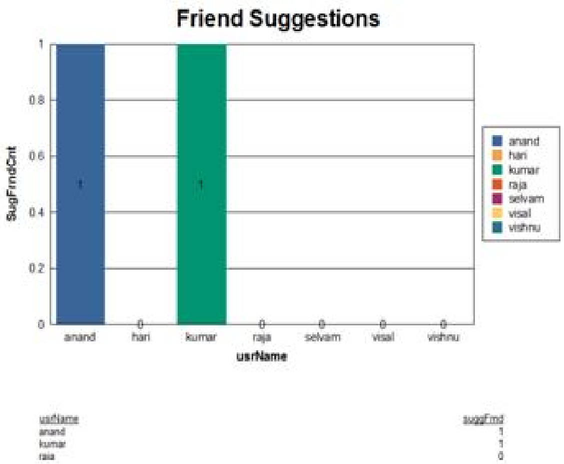


Figure5.2: Performance Evaluation of the Implemented Friend Suggestion Algorithms

In Figure 5.2 , we can able incur the results obtained by friend suggestion algorithm to the community user to enlarge the network also the technique is capable of detecting the fake or overlapping users in the community .

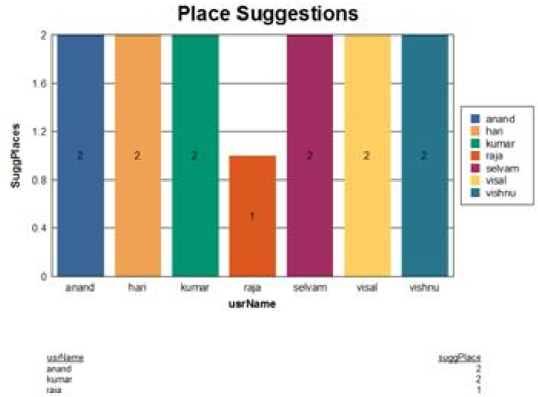


Figure5.3: Performance Evaluation of the Implemented Place Suggestion Algorithms

In Figure 5.3, we derived the outcome for community members for place recommendation based on the place familiarity and users place visit frequency in the particular community.

Table 5.1 : Performance analysis of the Community Discovery frameworks

Technique	Proposed	Existing (multicomm)
Precision	0.94	0.92
Recall	0.90	0.90
Fmeasure	0.92	0.91

The above table illustrate that proposed technique has some advantages over the other MultiComm algorithms that is one with no direct interaction between the same entities; (ii) the second is that the interactions are duplicated in the matrix form. Local modularity changes with respect to the number of items joined in the community on two generated multi-dimensional networks. As each of these two multi-dimensional networks is represented by multiple tensors, here the local modularity refers to the average value of local modularity's corresponding to these tensors.

VI. FUTURE ENHANCEMENTS

In the future work, it is required to adapt the proposed model to be time varying. As probability distributions are non-stationary in this situation, we must consider and study statistically dependence in time-varying Markov chains for items of different dimensions to obtain the affinities among them in order to find an evolution of communities across different time stamps.The major problem in Community Discovery is multidimensional context, and by introducing also a new measure able to characterize the communities found. We then provide a complete framework for finding and characterizing multidimensional communities

REFERENCES

[1] MAdFraud: Investigating Ad Fraud in Android Applications.

[2] Mining Personal Context-Aware Preferences for Mobile Users.

[3] A Flexible Generative Model for Preference Aggregation.

[4] Opinion spam and analysis. In Proceedings of the 2008 International Conference on Web Search and Networking, WSDM ’08, pages 219–230, 2008.

[5] D. M. Blei, A. Y. Ng, and M. I. Jordan. Lantent dirichlet allocation. *Journal of Machine Learning Research*, pages 993–1022, 2003.

[6] Y. Ge, H. Xiong, C. Liu, and Z.-H. Zhou. A taxi driving fraud detection system. In *Proceedings of the 2011 IEEE 11th International Conference Neural Networks*, ICNN ’11, pages 181–190, 2011.

[7] N. Spirin and J. Han, “Survey on web spam detection: Principles and algorithms,” SIGKDD Explor. Newslett., vol. 13, no. 2, pp. 50–64, May2012.

[8] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, “Detecting product review spammers using rating behaviors,” in Proc.19thACMInt. Conf. Inform. Knowl. Manage., 2010, pp. 939–948.

[9] W.Dzwinel et al Non multidimensional scaling and visualization of earth quake cluster over space and feature space, nonlinear processes in geophysics 12[2005] pp1-12.

[10] C.Lomnitz. Fundamentals of Earthquake prediction [1994].

[11] B.Gutenberg & C.H. Richtro, Earthquake magnitude, intensity, energy & acceleration bulseism soc. Am 36, 105-145 [1996],C.Brunk, J.Kelly & Rkohai “Mineset An integrate system for Visual Data Mining”

[12] http://www.dotnetspider.com.