iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 3 Issue: XII Month of publication: December 2015

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Volume 3 Issue XII, December 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

An Ensemble Approach to Discovering Evolving Communities in the Multidimensional Social Network

Ms. I. Ajitha[1], Mrs.M.Vasanthi[2]
[1]*Research Scholar,* [2]*Associate Professor*
Department of Computer Science, Sri Jayendra Saraswathi Arts and Science College

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. Unfortunately, the existing Community discover process-mining approaches do not take into account the hidden aspect 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. We assume that the intentional process models underlying user activities by using Intention mining techniques can discover the important information and entities as a community to gather the large members and exploit information's. The aim of this paper is to propose the use of probabilistic models to evaluate the most likely intentions behind traces of unobserved activities and mixture information's in the complex data structures between the multidimensional data, namely Discrete Hidden Markov Models (HMMs). Experiments based on synthetic and real-world data sets suggest that the proposed framework is able to find a community effectively. Experimental results have also shown that the performance of the proposed algorithm is better in accuracy than the other testing algorithms in finding communities in multi-dimensional networks
Keywords: Online Social Networks, Multidimensional Data mining, Hidden Markov Model

I. INTRODUCTION

Location Based Services growing interests in studying and analyzing large networks such as social networks, genetic networks and co-citation networks. In these networks, each node is an item corresponding to a dimension or an entity in a network. Each edge indicates a relationship between two nodes, for instance, a contact between two users in a social network, an interaction between two genes in a genetic network, and a citation between two papers or two authors in a co-citation network. Analyzing these networks enable us to understand their topological properties and structural organization, One of such objectives is to detect communities or modules in large networks. One approach is to partition the network into sub-networks so that nodes in each sub-network are densely connected while nodes in different subnetworks are loosely connected. In online social media, networked data consists of multiple dimensions/ entities containing tags, photos, comments and stories [10]. We are interested to find a group of users who interact significantly on these media entities. In a co-citation network, we are interested to find a group of authors who cite/collaborate to each other (or a set of papers which are related to each other) significantly on publication information in titles, abstracts, and keywords as multiple dimensions/entities in the network . A data mining technique is proposed to detect the frequently interacting users based on the common subjects and grouping them in single community. Unfortunately, the existing Community discover process-mining approaches do not take into account the hidden aspect 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. We assume that the intentional process models underlying user activities by using Intention mining techniques can discover the important information and entities as a community to gather the large members and exploit information's. The aim of this paper is to propose the use of probabilistic models to evaluate the most likely intentions behind traces of unobserved activities and mixture information's in the complex data structures between the multidimensional data, namely Discrete Hidden Markov Models (HMMs). Experiments based on synthetic and real-world data sets suggest that the proposed framework is able to find a community effectively. Experimental results have also shown that the performance of the proposed algorithm is better in accuracy than the other testing algorithms in finding communities in multi-dimensional networks. The rest of paper is organized as follows; section 2 explains the background knowledge regarding the related work. Section 3 explains and formulates the proposed System. The experimental results are discussed in section 4; we conclude the work with future work of the paper at section 5.

II. RELATED WORKS

A. A Local Method For Detecting Communities

A method of community detection based on supervised classifier 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.

B. Spectral Characterization And Scalable Mining Of Network Communities

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 can be inferred from their spectral signatures. Based on the above-mentioned relationship, this work proposes a general framework for characterizing, analyzing, and mining network communities. Utilizing the two basic properties of metastability, i.e., being locally uniform and temporarily fixed, an efficient implementation of the framework, called the LM algorithm, has been developed that can scalable mine communities hidden in large-scale networks.

C. MetaFac: Community Discovery Via Relational Hyper Graph Factorization

Discovering community structure in rich media social networks, through analysis of time-varying, multi-relational data is analysed. Community structure represents the *latent* social context of user actions. It has important applications in information tasks such as search and recommendation. Social media has several unique challenges. (a) In social media, the context of user actions is constantly changing and co-evolving; hence the social context contains time-evolving multi-dimensional relations. (b) The social context is determined by the available system features and is unique in each social media website. In this literature utilizes MetaFac (MetaGraph Factorization), a framework that extracts community structures from various social contexts and interactions. This work has three key contributions: (1) metagraph, a novel relational hypergraph representation for modeling multi-relational and multi-dimensional social data; (2) an efficient factorization method for community extraction on a given metagraph; (3) an on-line method to handle time-varying relations through incremental metagraph factorization.

D. MultiRank: Co-Ranking For Objects And Relations In Multi-Relational Data

We can explore a co-ranking scheme for objects and relations in multi-relational data. It has many important applications in data mining and information retrieval. However, in the technique, there is a lack of a general framework to deal with multi-relational data for co-ranking. The main contribution of this technique is to (i) propose a framework (MultiRank) to determine the importance of both objects and relations simultaneously based on a probability distribution computed from multi-relational data; (ii) show the existence and uniqueness of such probability distribution so that it can be used for co-ranking for objects and relations very effectively; and (iii) develop an efficient iterative algorithm to solve a set of tensor (multivariate polynomial) equations to obtain such probability distribution.

III. PROPOSED SYSTEM

A. Establishing Online Social Network Or Data Pre-Processing Of OSN dataset

In this Module, we either build the online Social network to construct the profile structure to yield the profile data with multidimensional fields which is considered to the synthesis dataset. Another way to gather is multidimensional data for a community discovery process is through real dataset from the Online Social network (OSN). Data Pre-processing involves stemming and Stop word removal process.

B. Partitioning Of The Network Using Spectral Methods

In this module, Social data network make use of the eigenvectors of matrix representations of the network. We show that with certain choices of the 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.

Apply Spectral similarity-based clustering to nodes

Vertex similarity is defined in terms of the similarity of their neighborhood

Structural equivalence: two nodes are structurally equivalent if they are connecting to the same set of actors

Optimal solution: top eigenvectors with the smallest eigenvalues

Structural equivalence is too restricting for practical use.

C. Establishing A MultiComm Community Discovery Model Based On Affinity Calculation

Data is partitioned as tensors .In this probability Distribution is used to calculate maximum likelihood between the data which affinity. 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. 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.

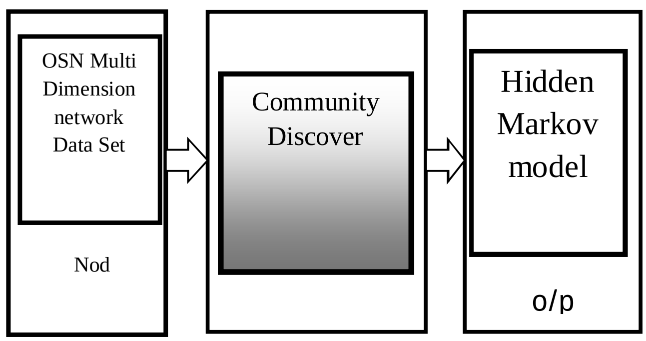


Figure 4.1 Architecture diagram of the community discovery framework

In figure 4.1, we explain the important components of the community discovery framework. 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. We will define the goodness criterion in order to determine the "best" community. Our idea is to 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. Based on this concept, we set the following tensor equations to calculate the required probabilities of visiting items in the vth dimension in the whole network.

Algorithm to Discovery the Community based Multi dimensional data using Hidden Markov Model

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

Value types = {Single Value Attribute, Two Value Attribute …. Multi- Value attributes}

Domain Knowledge = {Personal info, Employment, Lifestyle, Sports, Entertainments, cuisines}

Inference of the Data through Application is carried out

Application analysis is carried out for behaviours

Application Gain is calculated based on the attributes

Behaviours = Trained Data of the learning Algorithm

Behaviour = No. of Similarities between the attribute and data source

Affinity value is calculated based probability tensor

Probability is denoted P

$P= (1-\alpha)\ W_p +\alpha\ e_I$ -------> [1]

Above equation is a steady state probability

Where W is weighted matrix associated graph details of the User profiles using hidden Markov Model

Sensitivity is classified based on the application characteristic sand behaviour valuate

Largest value in the W and α leads to the maximum Support to form the Community group

Learning Algorithm extracts data based on the application category

Output: Secured disclosure of the information

In the algorithm, the computations require several iterations, through the collection to adjust approximate probability values of items of the entities in the multidimensional network to more closely reflect their theoretical true values. When communities vary in different subsets of dimensions, we can make use of affinity to identify which dimension of its corresponding item with the highest probability joins the community

D. Applying Feature Normalization And Fusion To Eliminate The Overlapping Profiles Using Hidden Markov Model

Due to the characteristic of various similarity features, different calculation methods might be used which lead to different value ranges. Therefore, the absolute values of different features must be normalized. Classical co clustering is one way to conduct this kind of community partitioning. Clauset defined a measure of community structure for a graph. Here we extend this idea to define a local modularity of a community in a multi-dimensional network. The identified communities are disjointed, which contradicts with the actual social setting. Edge clustering has been proposed to detect communities in an overlapping manner.

1) Normalized Mutual Information: Entropy: the information contained in a distribution
Mutual Information: the shared information between two distributions
Normalized Mutual Information (between 0 and 1)
Consider a partition as a distribution (probability of one node falling into one community), we can compute the matching between the clustering result and the ground truth

IV. EXPERIEMENTAL RESULTS

We have proposed a framework (Community discovery framework) to determine communities in a multi-dimensional network

Volume 3 Issue XII, December 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

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 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.

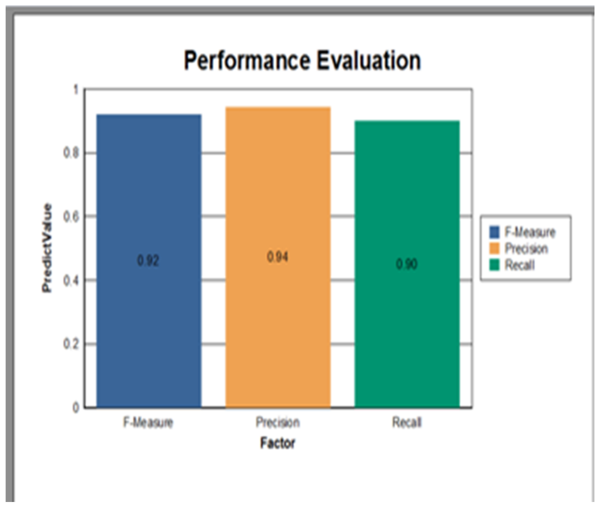


Figure 4.1: Performance Evaluation of the Implemented Community Prediction Algorithms

Experimental results showed in figure 4.1 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. 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 .

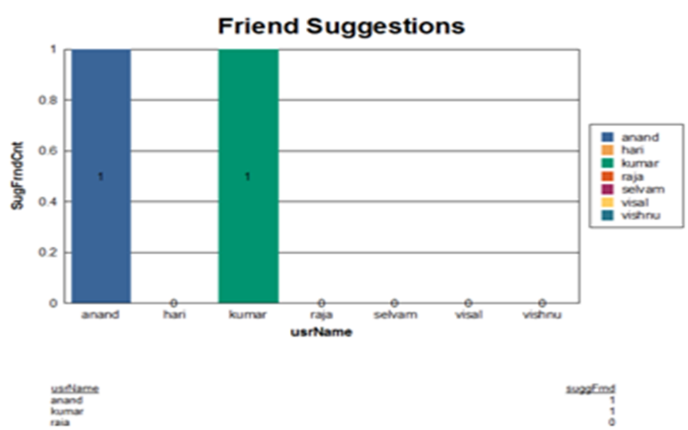


Figure5.2: Performance Evaluation of the Implemented Friend Suggestion Algorithms

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

Volume 3 Issue XII, December 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

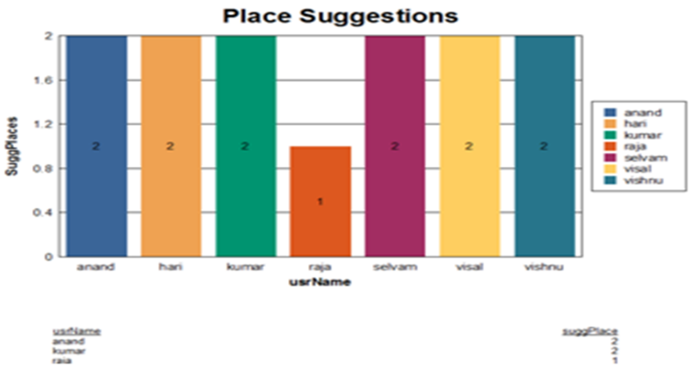


Figure5.3: Performance Evaluation of the Implemented Place Suggestion Algorithms

Table 5.1illustrate 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.

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

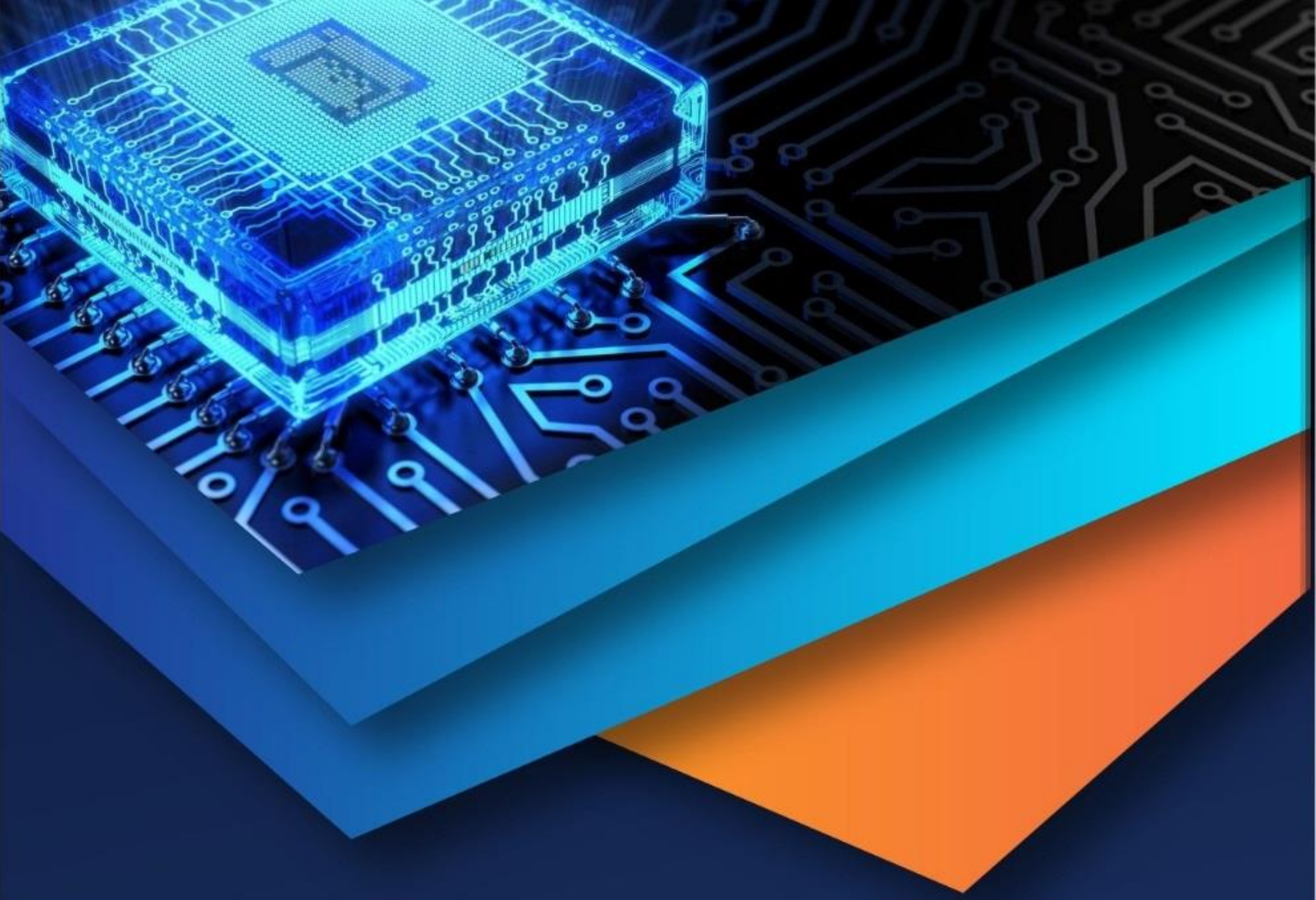
Local modularity changes with respect to the number of items joined in the community on two generated multi-dimensional networks. As each of these two multi-dimensional networks is represented by multiple tensors, here the local modularity refers to the average value of local modularity's corresponding to these communities.

V. CONCLUSION

In this Paper, we designed and implemented a framework named community discovery using discrete hidden markov model to discover the hidden community in the multidimensional network. It is 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. On the other hand, in social networks, user actions are constantly changing and co-evolving.

REFERENCES

[1] M. Girvan and M. Newman, "Community Structure in Social and Biological Networks," Proc. Nat'l Academy of Sciences USA, vol. 99, no. 12, p. 7821, 2002.
[2] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network Motifs: Simple Building Blocks of Complex Networks," Science, vol. 298, no. 5594, p. 824, 2002.
[3] M. Newman, "The Structure and Function of Complex Networks," SIAM Rev., vol. 45, no. 2, pp. 167-256, 2003.
[4] G. Palla, I. Der_enyi, I. Farkas, and T. Vicsek, "Uncovering the Overlapping Community Structure of Complex Networks in Nature and Society," Nature, vol. 435, no. 7043, pp. 814-818, 2005.
[5] S. Strogatz, "Exploring Complex Networks," Nature, vol. 410, no. 6825, pp. 268-276, 2001.
[6] M. Newman and M. Girvan, "Finding and Evaluating Community Structure in Networks," Physical Rev. E, vol. 69, no. 2, p. 026113, 2004.
[7] G. Flake, S. Lawrence, C. Giles, and F. Coetzee, "Self-Organization and Identification of Web Communities," Computer, vol. 35, no. 3, pp. 66-70, 2002.
[8] B. Yang, J. Liu, and J. Feng, "On the Spectral Characterization and Scalable Mining of Network Communities," IEEE Trans. Knowledge and Data Eng., vol. 24, no. 2, pp. 326-337, Feb. 2012.
[9] J. Ruan and W. Zhang, "An Efficient Spectral Algorithm for Network Community Discovery and Its Applications to Biological and Social Networks," Proc. Seventh IEEE Int'l Conf. Data Mining (ICDM '07), pp. 643-648, Jan. 2007.
[10] M. Shiga, I. Takigawa, and H. Mamitsuka, "A Spectral Clustering Approach to Optimally Combining Numericalvectors with a Modular Network," Proc. 13th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '07), pp. 647-656, 2007.
[11] S. Smyth, "A Spectral Clustering Approach to Finding Communities in Graphs," Proc. Fifth SIAM Int'l Conf. Data Mining (SDM '05), 2005.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429