A Perspective Analysis of Hidden Community Mining Methods in Large Scale Social Networks

Renuga Devi. R
Department of Computer Science,
Karpagam University
Coimbatore- 641 021

Hemalatha. M
Department of Computer Science,
Karpagam University,
Coimbatore – 641021

ABSTRACT
Research in social network analysis has increased in recent years. Because of the popularity of the social networking sites, many researchers concentrate on this area for research. In this, community mining plays an important role. Hidden communities affect the social networks in different ways. But not all hidden communities are dangerous or illegal. Most of the hidden communities are having potential knowledge. Communities are represented as a graph format. People are represented as nodes, and the relationship between the nodes are represented as edges. Several mining techniques do not considered the disconnected edges in the graph. Those hidden or disconnected edges may useful to the others in the network. Our approach on social network is fully based on the community mining on heterogeneous network. Here we analysis the various community mining techniques which is already available. Such as MinCut algorithm, Regression based algorithm, Max-Min modularity measure, LM algorithm and SECI model. Our results show that, there are some limitations in the hidden community mining technique in large scale networks. So we planned to do research in this area for better improvement.

Keywords
Social networks, Community Mining, Hidden Communities, Disconnected edges, mining techniques.

1. INTRODUCTION
Social networking sites are continually growing now a day, because, millions of users and the large amount of information shared in the sites. Social Networking Internet services are used to change the way of communication. Social Networking is one of the main reasons that many people have become passionate Internet users. The growing capacity of the Internet, community mining has much attention. Now a day there are so many numbers of research works has been in the Community mining. We can define a community as a group of objects. These objects may share some common information. Community mining Problem has some similar properties to the graph-cut problem, Kumar et al. In general both social network analysis and community mining has been seen as graph mining. The community mining is considered as a sub-graph identification [1]. Finding the hidden communities in the connection and focusing the entities in various communities has been important work of community mining for that it has various applications. Group of links within the connections are dense [2]. The Community mining problem treated as an application-oriented problem. Because the structure of the network is depends on the particular

applications. Several existing methods are based on the human observations. Bo yang et al., classified the existing methods into two main categories, explicit and not explicit optimization [3]. The explicit optimization techniques are used to solve the community mining problem by converting the problem into an optimization problem. And then it tries to find a solution for a predefined objective function, such as different spectral methods [4], [6], [7], [8]. The methods using not explicit optimization objectives are used to solve the community mining problem based on predefined assumptions or heuristic rules. Example for this method is maximum flow community (MFC) algorithm [3].

Existing data mining algorithms, like association rule mining, supervised classification algorithms and clustering algorithms are usually plan to notice patterns during a knowledge set characterized by a set of freelance instances, that is in line with the classical applied math illation drawback of making to spot a model given an independent, identically distributed (IID) sample [9]. However, a replacement rising challenge that data mining researchers face is resolution the community mining drawback on richly structured, heterogeneous knowledge sets.

Richard Ribeiro and Chris Kimble examine the ways in which to spot the 'hidden' community of Practice (CoP) within the network. COPs, is a neighbourhood of constant study and analysis (Lave, Wenger, McDermott, and Snyder) [10]. It is like, creating, functioning and management of information in social communities. It is associated with numerous places, like education, business, organizations and analysis industries. Communities of practices must have the following three characteristics. They are, Domain, Community and practice. In that, domain is a place for creating a general identity for the group members.

Bo yang et al., proposed the LM algorithm, to practically solve large-scale NCMPs. Mainly used for the community mining problem based on Local Mixing properties. In that they discussed about the relationship between the network structure and the Markov chains Meta stable states. They proposed measurement called spectral signature to characterize and analyze network communities, which is based on the striking and exiting times of states. Based on the network relationship and the spectral signature, they proposed a framework for characterize, analyze, and mine the network communities [3].

In this paper, we are taken four research papers for survey. These Papers are based on the hidden community mining problem. The paper is organized as follows. First section of the paper is Deng cai et al's, Mining Hidden Communities in Heterogeneous Social networks. Here we discussed about the points which is given by the Deng cai et al. Second section of the paper is Richard Ribeiro

and Chris kimble's, identifying hidden communities of practice. Third section is Jiyang Chen et al's., Detecting Communities in social networks using Max-Min Modularity. Fourth section is Bo Yang et al's., Spectral Characterization and Scalable Mining of Network Communities. Final section is discussion part and the conclusion.

2. MULTI RELATIONAL NETWORK

In a social network environment many kind of relationships are exist. Each relation can be treated as a relational network. Such kind of social network can be called multi relational and heterogeneous social network.

2.1 Mining Hidden Communities

Deng cai et al, proposed a method to mine the hidden communities in heterogeneous networks. Through their analysis, most of the existing methods focus on single relations in the networks, and provide results, which is independent of the user needs and references. In a single community mining method only one kind of relationship only considered. Remaining valuable relationships and their information might be missed. To overcome this problem they have taken the problem of mining hidden communities on heterogeneous network. Based on the study, observed that different relations may have different importance [11, 12]. So they proposed a method for learning optimal linear combination of these relations. That is used to fulfill the user expectations on relations. In the proposed method they shift the mining methods in advanced level. For their experiments, have taken two kinds of dataset, one is Iris dataset and DBLP dataset [1]. Each relation has modeled as a different graph. And each graph is considered as different communities. For relation extraction it can be non- linear or non linear. But in the case of real world application, it is difficult to collect the data. And non-linear techniques are unstable and cause over fitting problem. So they focused the linear techniques.

They define the graph as $G_i(V, E_i)$ where I = 1, 2, 3 …….. , n. and n is the number of relations, V is the set of nodes, E_i is the set of edges with respect to i-th relation. Here natural weights are assigned the based on the relationship strength of nodes. The M_i is used to denote the weight matrix, which is associated with G_i. If there is a hidden relation, then it can be represented as a graph $\hat{G}(V, \hat{E})$. Here $\hat{M}$ used to denote the weight matrix associated with the $\hat{G}$. They have taken set of labeled object for their experiments.

$$X = [x_1, x_2, x_3 \ldots x_m]$$
$$Y = [y_1, y_2, y_3 \ldots y_m]$$

Where y_j is the label of x_j. These labeled objects used to indicate the partial information of the hidden relation $\hat{G}$, which is used to find the linear combination of the weight matrixes. And it provides the better estimation of hidden matrix $\hat{M}$ [1].

2.1.1 A regression Based Algorithm

The authors used regression based algorithm to find a relation, which is a combined relation, contains the relationship between the intra community examples and the relationship between the inter community examples. For each relation was normalized it to make the biggest strength be 1.

So the target relation was constructed between labeled objects.

$$\tilde{M}_{ij} = \begin{cases} 1, & i \text{ and } j \text{ have the same label}; \\ 0, & otherwise. \end{cases}$$

Where $\tilde{M}$ is an m*m matrix. $\tilde{M}_{ij}$ indicates the relationship between the examples I and j.

They defined $\tilde{M}$ as $a^{opt} = \text{Prob}(x_i \text{ and } x_j)$. Both x_i and x_j belongs to the same community. This extraction relation problem is interpreted as a prediction problem.

After combining the combination coefficients, predict the hidden relation strength between the pair of objects. The objective function models the relation extraction problem as an unconstrained linear regression problem. Main advantages of linear unconstrained linear regression are, it has a close form solution and it is very easy to compute. But most of the times it might not be satisfactory. It requires coefficient shrinkage technique for two reasons. First one is Accuracy prediction and the other one is interpretation. In order to get the accuracy, it sacrifices a little bit bias value. For interpretation, it sacrifices some amount of information. Deng cai et al, called this method as ridge expression. And it can be solved by some numerical methods. But these regression methods may fail, when the examples provided by the user belong to only one community that is single community [1].

2.1.2 A MinCut – Based Algorithm

In order to deal with the single community problem, Deng cai et al, took the weakest connection in the extracted relation. To evaluate the tightness of the graph the minimum cut values are used on the graph. The cut graph was defined as a set of edges. These edges separate the vertices into two disconnected groups. That groups are denotes by A and B, where $A \cap B = \Phi$ and $A \cup B = G$. Here G is used to denote a weighted graph, M to denote the weight matrix, and m to denote the number of vertices. They took the cut value as

$$cut(G) = \sum_{i \in A} \sum_{j \in B} M(i, j).$$

And they defined the minimum cut as
$$\min cut(G) = \min_k \{cut_k(G)\}.$$

If a graph has a minimum cut value, then it is easy to cut into two sub graph. But in the case of disconnected graph its minimum cut value is 0. So in order to deal with the single community issue, the authors extract the optimal graph by maximizing its minimum cut value.
Their objective function is

$$a^{opt} = \arg\max_a \left\{ \min cut \left(\sum_{i=1}^{n} a_i G_i \right) \right\}$$

They have taken the small number of examples which is given by the user. So they have used linear programming technique to solve the optimization problem. This MinCut - Based algorithm is mainly used for the single community example problem [1].

3. IDENTIFYING HIDDEN COMMUNITIES OF PRACTICE

Richard Ribeiro and Chris kimble examine the possible ways to identify the hidden communities of practice (CoPs) in the electronic network. They mainly focused on the people in the hidden community who are all having the potential knowledge.

And they use this knowledge to develop a virtual community of practice (VCoPs). Questionary model are developed for the work. Questions related to the following topics. The characteristics of hidden Communities are the relation between Communities of practices, Virtual Communities, Distributed Community of practices and Virtual Community of practices (VCoPs). Interviews are conducted for the research in the higher education academy psychology network (UK) [13].

3.1 Model for knowledge transformation

CoP brings novel ways to manage the knowledge. And then this knowledge is shared and is formed by the sum of individual and internal knowledge. This individual knowledge is known as tacit knowledge. A Person may have two types of knowledge. Those are tacit knowledge and explicit knowledge. Tacit knowledge is very hard to acquire and fully personal. In the early stages of 1990s, based on the Nonaka suggestions, this tacit knowledge is most useful for success of many companies. He proposed SECI Model for the use of tacit knowledge, but still few authors argue about this SECI Model. For example, Gourlay told that, the SECI model not supported by empirical evidence and some of its phases are not coherent. Jorna told that, the SECI model lacks of background in learning theories and lacks in methodology [5].

3.2 SECI Model

SECI model based upon the idea of apprenticeship. It has four phases to transfer tacit knowledge. Each phase represents unique movement between tacit and explicit. Nonaka represent these phases as way of knowledge conversion. The SECI model is also known as spiral of knowledge or knowledge spiral. Four phases of SECI model is, Socializations (S), Externalizations (E), Combination (C), Internalizations (I). In that Socialisation phase, the apprentice acquires the basic skills from their co workers. In the second phase Externalisation, after acquiring the tacit knowledge transfer it to a media or pass it on. In the third phase Combination, explicit knowledge is combined with existing explicit knowledge. The last phase is internalization, where the knowledge after the interactions phase becomes novel, wealthier and more expanded, tacit knowledge. According to the Nonaka prediction this SECI model is more useful for distributed environment [5].

3.3 Hidden CoPs

We can utilize the benefits of CoPs only when the community is fully working. But in few cases, the CoPs are unknown or not established yet. And in few times these communities are 'hidden', or potential communities. The term 'hidden' does not meant that illegal or dangerous, it simply states that the CoPs are not fully visible or developed.

The term 'Hidden' CoPs are not new. In the year of 1991, Brown and Duguid used this term. In 1998, Wenger used, and again in the year of 2002, Wenger et al., also used this hidden CoPs. They called these hidden communities as loose networks. Richard Ribeiro and Chris kimble studied some basic concepts of CoPs before entering into their process, that are, Distributed Communities of Practice (DCoPs), Virtual Communities (VC), and Virtual Communities of Practice (VCoPs) [5].

3.4 Fully Developed Virtual Communities of Practices

A well developed VCoPs must have any one of the following characteristics, Such as Social Community, CoP, DCoP and VC. Developing a well defined VCoPs is fully depending on the

discovery of hidden Communities. For that Richard Ribeiro and Chris kimble analysis some issues [5].

1) Difficult to discover the Communities desires. Original CoPs cannot be created by force or artificially.
2) Analysis on a case to case basis (Interviews)
3) Time for an intervention to keep the VCoPs active [13].

4 FINDING HIDDEN COMMUNITIES IN SOCIAL NETWORKS USING MAX-MIN MODULARITY

Jiyang Chen et al., Prepared a new mining method called Max-Min Modularity. It considers both connected pairs and criteria defined by domain experts in community mining. And they specify a hierarchical clustering algorithm to detect communities in network. They applied these methods to both real world networks and randomly generated networks. For real world applications it gives better results when compared to previous algorithms. And for randomly generated networks it gives promising results. Different relation-based methods have been developed. The spectral clustering methods [14, 15, and 16] and modularity based algorithms [17, 18]. However, none of them differentiate the essential features of the area of the network in question. Jiyang Chen et al., presented a novel modularity-based measure called Max-Min Modularity. It considers the connected pairs and user-defined related node pairs in the community finding.

There are many graph partitioning algorithms and hierarchical clustering algorithms are available. Existing algorithm are failed in unknown networks and the structureless networks. To solve this problem, computer scientists proposed several benefit functions based on cut sizes, e.g., normalized cut [15]. So they used modularity measure.

4.1 Modularity measure

Proposed by Newman and Girvan [19]. It is a measure of the quality of a part of the network. And they defined the modularity measure as

Q= (Number of edges within communities) – (expected number of edges)

Their main idea is comparing the division with a null model, randomly generated networks same to the network.

$$Q = \sum e_{II} - a_i^2) = T_v(e) - ||e^2||$$

Where e= Symmetric matrix k*k. k division of a network can be represented as k * K. The matrix trace ||x|| indicates the sum of the elements of the matrix x. for the number of edges between groups is smaller than expected.

4.2 Drawbacks

It t has three major drawbacks.
1)   Modularity measure requires the entire structure information. This is problem for huge network. To overcome this problem Clauset used a measure of local community measure, called Local modularity. But there is lack of global knowledge [19].
2)   Modularity measure have resolution limit and fails to identify communities smaller than a certain scale [20]. To solve this problem, they used recursive algorithms based on modularity optimization [21].
3)   Modularity measures observe only the existing links, does not consider absent links. Si it fails to compare community structure between different links. For this Scripps proposed two ratios p and q fro absent links [22].

So Jiyang Chen et al., proposed a new measure for mining. They separate the disconnected pair set into two categories based on the domain experts. First one is, the related pair set. And the second category is unrelated pair set. Here the nodes are certainly unrelated. They penalized their measure score if they identify unrelated pairs based on the Max-Min (MM) modularity. MM is not only based on the division of a network into community is not merely one in which the number of edges between groups is smaller than expected. And also the number of unrelated pairs in with in groups is smaller than expected, if both are significantly lower than expected one. Their aim is to increase the number of edges within groups and decrease the number of unrelated pairs from user defined unrelated pair set. So they name their measure as Max-Min.

4.3 Generalization of the Max-Min Modularity

The modularity Q was transformed from community based to node based. They defined an unweighted and undirected network as $G = (V, E)$. They denote the A_{xy} as an element of the adjacency matrix of G.

$$A_{xy} = \begin{cases} 1 \ if \ vertices \ x \ and \ y \ are \ connected \\ 0, otherwise \end{cases}$$

Then the network divided into K-Communities. Node x belongs to C_x community, and fraction of edges fall between community I and j.

Then they got the modularity Q as following

$$Q_{\max} = \frac{1}{2m} \sum_{xy} [A_{xy} - P_{xy}] \Phi(C_x, C_y)$$

If the values of Q is less than it is provides better communication division. And this Max-Min modularity does not require extra computation or materialization. Just we have to maintain and update the related pair set defined by the domain experts.

They defined used defined criteria U to define two disconnected nodes i, j are related to $(i, j) \in U$ or unrelated $(i, j) \notin U$.

So they create

$G' = (V, E')$. Here G' is the complement of G graph, and A' is the adjacency matrix of G'.

$A'_{ij} = 1$ if $A_{ij} = 1$ and $(i, j) \notin U$.

$\Phi(C_x, C_y) = 1$ if C_x, C_y is same community, or else zero. And $|U| =$ number of pairs in U.

P'_{xy} Is the predictable probability of an edge among vertices x and y in random graph.

$$P'_{xy} = \frac{(d'_x)(d'_y)}{2m'}$$ or

$$P'_{xy} = \frac{(n - d_x - u_x)(n - d_y - u_y)}{n(n-1) - 2m - 2|u|}$$

Here u_x, u_y are disconnected vertices abut related to graph. They provide a formulae for maximize the $Q_{\max}$ and minimize the $Q_{\min}$ at the similar time. That is

$$Q_{\max-\min} = Q_{\max} - Q_{\min}$$

If the $Q_{\max-\min}$ is higher, then it is provides better results. Even though this Max-Min Modularity provides better results, it has few disadvantages also. Such as, $Q_{\max-\min}$ is very costly to carry out a complete search all possible division. So they proposed $H_{\max-\min}$ algorithm.

The ARI (Adjusted Rand Index) used as the performance metric for accuracy. This ARI Measure used to measure the similar partition of objects according to the real community (R) and the partition in an algorithm result (P). They have taken four object pairs like a, b, c and d. They may belongs to the same community R and P, or in the R but not in the P, or both are belongs to different community R and P.

They defined ARI Index as follows

$$ARI(R, P) = \frac{2(a*d - b*c)}{(a+b)*(b+d) + (a+c)*(c+d)}$$. They used

ARI measure to test the randomly generated graph. That random graph contains 1000 vertices and five communities. They found that their $H_{\max-\min}$ algorithm provide better result of a>0.7 average ARI on graphs. All algorithms are working same if both communities are clear and strong. But if the graph contains noise edges then $H_{\max-\min}$ provides better results. They test their algorithms in the karate club [17], sawmill communication network, and Mexican politician network. This Max-Min modularity measures solves the problem of Newman's method. It extends local information to global level. And this Max-Min used in recursive communication detection. The Max-Min modularity measure extract gets information from links prediction, and extract appropriate criteria for community detection [23].

5. LM ALGORITHM

LM(Local Mixing properties) algorithm was developed by Bo Yang et al., mainly used to mine the hidden communities in large networks. The authors proposed a new measurement called spectral signature, to analyze and characterize the network communities. In this process, the stochastic model dynamics reaches the local mixing states before enter into the global state. For each transition distribution values are measured. Topological values of the networks are observed based on the entering and exiting time of the local state based on the large deviation theory. But it is difficult to find the implicit information of a network. They denote the network as $N = (V, E)$. Here V denotes the set of nodes and the E is the relationships or links of a network. In this process an agent can freely move from one node to another node along with the links. After reaching the first node, it randomly chooses its neighbor node, and makes a move on it. They denote the agent position as $\lambda = \{X_t, t \geq 0\}$. They found the probability value of the agent moves based on the steps to reach i to j node $p\{X_t = i, 1 \leq i \leq n\}$. They characterize the community as two community and k-community structure. In the two community network, the markov chain entered in to the one local state before entering into the global state [24]. If the both the communities are closer then the chain will reach the local state soon. Other it will more time to reach the state. The process will become slow. After entering into the local state the two network communities are separated and it take much time to exit the current state and mingle each other.

Based on the entering and mixing time of the local state the cohesion (C2) and separability (S2) values are measured.

$$C_2 = T_2^{hit} = \frac{1}{\lambda_3} \quad And \quad S_2 = T_n^{ext} - T_2^{hit} = \frac{1}{\lambda_2} - \frac{1}{\lambda_3}.$$

To characterize the k-community structure they denote the cohesion (C2) and separability (S2) as follows.

$$C_k = T_k^{hit} = \frac{1}{\lambda_{k+1}} \quad And \quad S_k = T_k^{ext} - T_k^{hit} = \frac{1}{\lambda_k} - \frac{1}{\lambda_{k+1}}.$$ They implemented the algorithm to cluster the entire network for

calculating and extracting a information from a single column distribution not to deal with the entire matrix. They implemented as follows. First they select the single column and then they distribute ordering time, and finally inferring network communities from ordering time distribution.

6. DISCUSSIONS

Deng cai et al, used Iris dataset and DBLP dataset. In their analysis they extract the optimal relation based on examples. They believe that, if the label information is available, then it is better to extract the optimal relation. In the Iris dataset, they have taken three classes, each class has 50 instances. Each instance has 4 main features, taken f3 and f4 as an important features. And construct four matrices M1, M2, M3, and M4 from the four features. In the DBLP (Digital Bibliography and Library project) dataset, they used author as a objects. It provides data in the XML format. They took author name, paper and conference details for their analysis.

Richard Ribeiro and Chris kimble prepared a case study for their work. They did analysis in Higher Education Academic Psychology Network (UK). This institution contains 24 departments. They formed 14 statements for verify the existence of three main components of a CoPs. That are, Mutual Engagement, a Joint enterprise and a share repertoire. They collect the answers from the participants. From the analysis identified that, the most of the participants have the beginning of a collocated CoPs. So, final study was needed to the use of CMC in CoPs.

Jiyang Chen et al., carry out the experiments on real time social networks. The Proposed algorithm provides better scalability. To check the scalability of the system, ran the algorithms in the largest components of the collaboration network of scientist posting preprints at WWW.arxiv.org [25]. Network has 27,519 nodes and 116,181 edges. And also test their algorithm in IMDB network. It has 47,436 nodes and 376,196 edges with in 376 and 1037 seconds. They also test their algorithms in randomly generated graphs. Graph contains the vertices ranges from 10000 to 500000 and edges ranges from 20,000 to 1,000,000. Algorithm was compared with the existing algorithms. The algorithm provides better result of a>0.7 average ARI on graphs. They got accurate results for community detection. But the system has few drawbacks. They did not considered the duplicate author name in the DBLP dataset. And DBLP does not provide the proceeding information for all papers. They used 'key' attribute with 'year' attribute. But this may not true for all cases [22].

Bo yang et al., compared the proposed LM algorithm with the already existing algorithms. They are GN algorithm [26], FN algorithm [27], Ncut algorithm [28] and Mincut algorithm [29], and the GA algorithm [30]. They tested it on the real world networks. Such as dolphin network [31], karate club [32], foot ball network [33], semantic network [34] and some other real world networks. Finally they stated that, their work is completely different from the existing work based on the three main aspects. First one is spectral graph division methods are basically optimization based methods. Second thing is, several existing spectral methods operate the connections among networks' eigenvectors values and their best partitions, but they have given the more effort to find the hidden communities in the connections. Final difference is the rationales behind spectral strategies and therefore the LM square measure fully distinct from the perspective of sensible calculation. The proposed LM algorithm is much efficient than the existing spectral methods based algorithm. But when compared the effectiveness of the LM algorithm, existing methods provide better results in the predefined networks but LM works better in real world

networks [3]. It has few limitations. They not considered the arbitrary networks. If the hitting time is low then only efficiency of the LM algorithm will be good. Otherwise, the efficiency is slow.

7. CONCLUSIONS

The existing approaches in social networking is just shift the networks from the traditional single network, user independent analysis to multi network user dependent and query based analysis. From our analysis, conclude that, electronic network, like Internet alone can able to create a new opportunities for practical uses of such communities in day to day life. But still studies are required, if VCoPs have similar behaviour to collocated ones. Hidden communities are still understudied. More case studies are required to develop a VCoPs. More analysis is necessary to make more interviews in different organization with a larger number of participants. All algorithms are working same if both communities are clear and strong. But if the graph contains noise edges then provides better results. Max-Min modularity solves the problem of Newman's method. It extends local information to global level. And this Max-Min used in recursive communication detection. LM algorithm detects the number of communities by using the recursive bisection method. And the stopping criteria are predefined. From our analysis, we planned to improve the existing algorithms without predefined values. And we planned to propose a new method for clustering using KD-Tree. Our research is fully based on the problems in the hidden community mining.

8. REFERENCES

[1] Deng Cai, Zheng Shao, Xiaofei He, Xifeng Yan, and Jiawei Han, *Mining Hidden Community in Heterogeneous Social Networks*, Report No. UIUCDCS-R-2005-2538 UILU-ENG-1731, March 2005.

[2] Osmar R. Za¨ıane, Jiyang Chen, and Randy Goebel, *Mining Research Communities in Bibliographical Data* , University of Alberta, Canada.

[3] Bo Yang, Jiming Liu, and Jianfeng Feng, *On the Spectral Characterization and Scalable Mining of Network Communities*. IEEE Transactions on Knowledge and Data Engineering. VOL. 24, No. 2, pp. 1041-4347, 2012.

[4] M. Fiedler, *Algebraic Connectivity of Graphs*, Czechoslovakian Math. J., vol. 23, pp. 298-305, 1973.

[5] Richard Ribeiro and chris kimble,. *Identifying 'Hidden' Communities of Practice Within Electronic Networks: Some Preliminary Premises*, In Proceedings of 13th UKAIS Conference, Bournemouth, UK, 2008

[6] M.E.J. Newman, *Modularity and Community Structure in Networks*, Proc. Nat'l Academy of Sciences USA, vol. 103, no. 23, pp. 8577-8582, 2006.

[7] S. White and P. Smyth, *A Spectral Clustering Approach to Finding Communities in Graphs*, Proceeding of Fifth SIAM Int'l Conf. Data Mining, 2005.

[8] M. Shiga, I. Takigawa, and H. Mamitsuka, *A Spectral Clustering Approach to Optimally Combining Numerical Vectors with a Modular Network*, Proc. 13th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 647-656, 2007.

[9] L. Getoor and C. P. Diehl. *Link mining: a survey*. SIGKDD Explor. Newsl., 7(2):3-12, 2005.

[10] Lave, J and Wenger. E. *Situated learning:Legitimate peripheral participation*, Cambridge University Press, 1991.

[11] M. F. Schwartz and D. C. M. Wood, *Discovering shared interests using graph analysis*. Communications of the ACM, 36(8):78–89, 1993.

[12] P. Domingos and M. Richardson, *Mining the network value of customers*. In Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, pages 57–66. ACM Press, 2001.

[13] Richard Ribeiro and Chris Kimble, *Identifying 'Hidden' Communities of Practice with in electronic networks: Some preliminary premises,* 13th UKAIS conference (UKAIS). 2008.

[14] C. H. Q. Ding, X. He, H. Zha, M. Gu, and H. D. Simon, *A min-max cut algorithm for graph partitioning and data clustering*. In ICDM, pages 107-114, 2001.

[15] J. Shi and J. Malik, *Normalized cuts and image segmentation*. IEEE. Trans. on Pattern Analysis and Machine Intelligence, 2000.

[16] S. White and P. Smyth. *A spectral clustering approach to finding communities in graphs*. In SIAM, 2005.

[17] M. E. J. Newman and M. Girvan. *Finding and evaluating community structure in networks*. Physical Review E, 69, 2004.

[18] A. Clauset. *Finding local community structure in networks*. Physical Review E, 72:026132, 2005.

[19] S. Fortunato and M. Barthelemy. *Resolution limit in community detection*. PROC.NATL.ACAD.SCI.USA,104:36, 2007.

[20] J. Ruan and W. Zhang. *Identifying network communities with a high resolution*. Physical Review E, 77:016104, 2008.

[21] J. Scripps, P.-N. Tan, and A.-H. Esfahanian. , *Exploration of link structure and community-based node roles in network*. In ICDM, 2007.

[22] W. W. Zachary. *An information ow model for conict and _ssion in small groups*. Journal of Anthropological Research, 33:452-473, 1977.

[23] Jiyang Chen, Osmar R.Zaiane and Randy Geobel, *Detecting Communities in Social Networks uaing Max-Min Modularity*. SIAM, 2009.

[24] B. Yang, W.K. Cheung, and J. Liu, *Community Mining from Signed Social Networks*, IEEE Trans. Knowledge and Data Eng., vol. 19, no. 10, pp. 1333-1348, Oct. 2007.

[25] M. E. J. Newman, *Fast algorithm for detecting community structure in networks*. Physical Review E, 69, 2004.

[26] M. Girvan and M.E.J. Newman, *Community Structure in Social and Biological Networks*, Proc. Nat'l Academy of Sciences USA, vol. 9, no. 12, pp. 7821-7826, 2002.

[27] M.E.J. Newman, *Fast Algorithm for Detecting Community Structure in Networks*, Physical Rev. E, vol. 69, no. 6, p. 066133, 2004.

[28] J. Shi and J. Malik, *Normalized Cuts and Image Segmentation, IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888-904, 2000.

[29] M. Fiedler, *Algebraic Connectivity of Graphs*," Czechoslovakian Math. J., vol. 23, pp. 298-305, 1973."

[30] R. Guimera and L.A.N. Amaral, *Functional Cartography of Complex Metabolic Networks,* Nature, vol. 433, no. 2, pp. 895-900, 2005.

[31] D. Lusseau,. *The Emergent Properties of a Dolphin Social Network*, Proc. Royal Soc. B: Biological Sciences, vol. 270, no. Suppl 2, pp. S186-S188, 2003

[32] W.W. Zachary,"*An Information Flow Model for Conflict and Fission in Small Groups,* J. Anthropological Research, vol. 33, pp. 452-473, 1977.

[33] M. Girvan and M.E.J. Newman, "*Community Structure in Social and Biological Networks,*" Proc. Nat'l Academy of Sciences USA, vol. 9, no. 12, pp. 7821-7826, 2002.

[34] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek, *Uncovering the Overlapping Community Structures of Complex Networks in Nature and Society*, Nature, vol. 435, no. 7043, pp. 814-818, 2005.

[35] M. Girvan and M. E. J. Newman, *Community structure in social and biological networks*. In PNAS USA, 99:8271-8276, 2002.

[36] M. E. J. Newman, *Fast algorithm for detecting community structure in networks*. Physical Review E, 69, 2004.