Journal of Mathematics and Statistics 4 (1): 51-57, 2008
ISSN 1549-3644

Framework of Regression-Based Graph Matrix Analysis Approach in Multi-Relational Social Network Problem

Ford Lumban Gaol and Belawati Widjaja
Faculty of Computer Science, University of Indonesia, Indonesia

Abstract: Community mining is one of the major directions in social network analysis. Social network analysis has attracted much attention in recent years. Most of the existing methods on community mining assume that there is only one kind of relation in the network and moreover, the mining results are independent of the users' needs or preferences. However, in reality, there exist multiple, heterogeneous social networks, each representing a particular kind of relationship and each kind of relationship may play a distinct role in a particular task. Thus mining networks by assuming only one kind of relation may miss a lot of valuable hidden community information and may not be adaptable to the diverse information needs from different users. In this research, we systematically analyze the problem of mining hidden communities on heterogeneous social networks. Based on the observation that different relations have different importance with respect to a certain query, we propose a method for learning an optimal linear combination of these relations which can best meet the user's expectation. With the obtained relation, better performance can be achieved for community mining.

Key words: Community mining, social network analysis, hidden community information

INTRODUCTION

Social Network Analysis (SNA) is the mapping and measuring of relationships and flows between people, groups, organizations, computers, or other information/knowledge processing objects. Social network analysis as a theme has been studied for years. The classic paper of Milgram[23] might be one of the first works on SNA. It estimates that every person in the world is only six edges away from every other. It sets the stage for investigations into social networks and algorithmic aspects of social networks. Many recent efforts try to leverage social networks for diverse purposes, such as expertise location[19,19], mining the network value of customers[11] and discovering shared interests[26].

Previous work in sociology and statistics has suffered from the lack of data and focused on very small networks, typically in the tens of individuals[29]. With the web growing, much potential social network data are available and a lot research efforts have been put on dealing with such data.

Schwartz and Wood mined social relationships from email logs[31]. The ReferralWeb project[19] is proposed to mine a social network from a wide variety of web data and use it to help individuals find experts who could answer their questions. Adamic and Adar tried to discover the social interactions between people from the information on their homepages[1].

Agrawal *et al.* analyzed the social behavior of the people on the newsgroups[2]. Moreover, the web itself can be acually viewed as a large social network.
The well-known link analysis algorithms, such as Google's PageRank[24,16,5,6] and Kleigberg's HITS algorithm[20], can be seen as social network analysis on the web.

Community mining: With the growth of the web, community mining has attracted increasing attention. A lot of work has been done at mining the implicit communities of web pages[14,22,7,12,27,13], scientific literature from the Web [31] and document citation database[25].

In principle, a community can be simply defined as a group of objects sharing some common properties. Community mining has many similar properties to the graph-cut problem. Kumar et al. used the bipartite graph concept to find the core of the community and then expanded the core to get the desired community[22]. Flake et al. applied the maximum-flow and minimumcut framework on the community mining[12]. The authority-and-hub idea[20] was also used in the community mining[14,21,8] and has several extensions[9]. The idea of frequent itemset in association rule mining has also been used in community mining[37].

Corresponding Author: Ford Lumban Gaol, Faculty of Computer Science University of Indonesia, Depok, Jawa Barat, Indonesia Tel: +62217863419 Fax: +62217863415

Generally speaking, both social network analysis and community mining can be seen as graph mining. The community mining can be thought of as sub-graph identification. Previous work on graph mining can be found in[10,28,30]. Almost all the previous techniques on graph mining and community mining are based on a homogenous graph, i.e., there is only one kind of relationship between the objects. However, in real social networks, there are always various kinds of relationships between the objects. To deal with this problem, we focus in this research on multi-relational community mining.

We begin on next section about relation extraction and follow with discussion about multi-relational social network and close with conclusion.

RELATION EXTRACTION

Here we begin with a detailed analysis of the relation extraction problem followed by two algorithms for two cases.

The problem: A typical social network likely contains multiple relations. Different relations can be modelled by different graphs. These different graphs reflect the relationship of the objects from different views.

For the problems of community mining, these different relation graphs can provide us with different communities.

In multi-relational social network, community mining should be dependent on the user's example (or information need). A user's query can be very flexible. Since previous community mining techniques only focus on single relational network and are independent of the user's query, they cannot cope with such a complex situation.

In this research, we focus on the relation extraction problem in multi-relational social network.

The community mining based on the extracted relation graph is more likely to meet the user's information need. For relation extraction, it can be either linear or nonlinear. Due to the consideration that in real world applications it is almost impossible for a user to provide sufficient information, nonlinear techniques tend to be unstable and may cause over-fitting problems. Therefore, here we only focus on linear techniques.

This problem of relation extraction can be mathematically defined as follows. Given a set of objects and a set of relations which can be represented by a set of graphs $G_i(V,E_i)$, i = 1, . . . , n, where n is the number of relations, V is the set of nodes (objects) and E_i is the set of edge with respect to the i-th relation. The weights on the edges can be naturally defined according to the relation strength of two objects. We use Mi to denote the weight matrix associated with G_i, i = 1,...,n. Suppose there exists a hidden relation represented by a graph $\hat{G}(V,\hat{E})$, and $\hat{M}$ denotes the weight matrix associated with $\hat{G}$. Given a set of labeled objects X = $[x_1,\ldots,x_m]$ and y = $[y_1,\ldots,y_m]$ where y_j is the label of x_j (Such labeled objects indicate partial information of the hidden relation $\hat{G}$), find a linear combination of the weight matrices which can give the best estimation of the hidden matrix $\hat{M}$.

A regression-based algorithm: The basic idea of our algorithm is trying to find an combined relation which makes the relationship between the intra-community examples as tight as possible and at the same time the relationship between the inter-community examples as loose as possible.

For each relation, we can normalize it to make the biggest strength (weight on the edge) be 1.

Thus we construct the target relation between the labeled objects as follows:

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

where, $\tilde{M}$ is a m×m matrix and $\tilde{M}_{ij}$ indicates the relationship between examples i and j. Once the target relation matrix is built, we aim at finding a linear combination of the existing relations to optimally approximate the target relation in the sense of L_2 norm. Sometimes, a user is uncertain if two objects belong to the same community and can only provide the possibility that two objects belong to the same community. In such case, we can define as follows $\tilde{M}$.

$\tilde{M}_{ij}$ = Prob (X_i and X_j belong to the same community)

Let $a = [a_1, a_2,\ldots, a_n]^T \in R^n$ denote the combination coefficients for different relations. The approximation problem can be characterized by solving the following optimization problem:

$$a^{opt} = \arg\min_{a} \left\| \tilde{M} - \sum_{i=1}^{n} a_i M_i \right\|^2 \qquad (1)$$

This can be written as a vector form. Since the matrix $M_{m\times m}$ is symmetric, we can use a m(m-1)/2 dimensional vector v to represent it. The problem (1) is equivalent to:

$$a^{opt} = \arg\min_{a} \left\| \tilde{V} - \sum_{i=1}^{n} a_i V_i \right\|^2 \quad (2)$$

Equation 2 is actually a linear regression problem[15]. From this point of view, the relation extraction problem is interpreted as a prediction problem. Once the combination coefficients are computed, the hidden relation strength between any object pair can be predicted.

In real applications, the user does not need to specify the relationships between any pair of objects. That is, the vector v need not to be m(m-1)/2 dimensional. We assume that v is a k-dimensional vector in the following.

Let us first consider the simplest case that:

$$\sum_{i=1}^{n} a_i v_i = \tilde{v} \quad (3)$$

We define:

$$V = [v_1, v_2, \ldots, v_n] \quad (4)$$

Thus, Eq. 3 can be rewritten as follows:

$$Va = \tilde{v}$$

Suppose the rank of V is min(k, n). We have the following facts:

- When k<n, the set of solutions to Eq. 4 forms a (n-k) dimensional vector subspace
- When k = n, there is a unique solution to Eq. 4
- When k>n, there is no solution to Eq. 4

In the first two cases, we get a solution with perfect match (The minimization error is zero). Note that, the value of k reflects the quantity of the user's information needs. k is small when the query submitted by the user is simple.

With a complex query, k can be larger than n. In this case, the optimal solution to (2) is obtained when the derivative of this objective function with respect to a is zero, i.e.

$$\frac{\partial \left\| \tilde{v} - \sum_{i=1}^{n} a_i v_i \right\|^2}{\partial a_i} = 0 \text{ for } i = 1, 2, \ldots, n$$

By some algebraic steps, we have:

$$\frac{\partial \left\| \tilde{v} - \sum_{i=1}^{n} a_i v_i \right\|^2}{\partial a_i} = 0$$

$$\frac{\partial \left[\left(\tilde{v} - \sum_{i=1}^{n} a_i v_i \right)^T \left(\tilde{v} - \sum_{i=1}^{n} a_i v_i \right) \right]}{\partial a_i} = 0$$

$$\frac{\partial \left(\tilde{v} - \sum_{i=1}^{n} a_i v_i \right)^T}{\partial a_i} \left(\tilde{v} - \sum_{i=1}^{n} a_i v_i \right) + \left(\tilde{v} - \sum_{i=1}^{n} a_i v_i \right)^T \frac{\partial \left(\tilde{v} - \sum_{i=1}^{n} a_i v_i \right)}{\partial a_i} = 0$$

$$2 \left\{ \frac{\partial \left(\tilde{v} - \sum_{i=1}^{n} a_i v_i \right)}{\partial a_i} \right\}^2 \left(\tilde{v} - \sum_{i=1}^{n} a_i v_i \right) = 0$$

$$v_i^T \left(\tilde{v} - \sum_{i=1}^{n} a_i v_i \right) = 0 \text{ for } i = 1, 2, \ldots, n.$$

$$v^T \left(\tilde{v} - \sum_{i=1}^{n} a_i v_i \right) = 0$$

$$v^T \left(\tilde{v} - Va \right) = 0$$

$$v^T Va = V^T \tilde{v}$$

Since the matrix V has full rank as we assumed, i.e., rank (V) = min (k, n), the matrix V^TV is invertible and the optimal solution to (2) is $a^{opt} = (V^TV)^{-1}V^T \tilde{v}$.

When the matrix V is rank deficiency, i.e., rank(V) < min(k, n), there will be multiple solutions with the same minimization value. In such case, we can choose the a with minimum norm as our solution[4].

The objective function (2) models the relation extraction problem as an unconstrained linear regression problem. One of the advantages of the unconstrained linear regression is that, it has a close form solution and is easy to compute. However, researches on linear regression problem show that in many cases, such unconstrained least squares solution might not be a satisfactory solution and the coefficient shrinkage technique should be applied based on the following two reasons[15].

Prediction accuracy: The least-squares estimates often have low bias but large variance[15].

The overall relationship prediction accuracy can sometimes be improved by shrinking or setting some coefficients to zero. By doing so we sacrifice a little bit of bias to reduce the variance of the predicted relation strength and hence may improve the overall relationship prediction accuracy.

Interpretation: With a large number of explicit (base) relation matrices and corresponding coefficients, we often would like to determine a smaller subset that exhibit the strongest effects.

In order to get the big picture, we are willing to sacrifice some of the small details.

Such consideration can be shown in the following example. Suppose we have a user query (o_1, o_2, o_3, o_4, o_5), where o_1, o_2 and o_3 belong to one community, but o_4 and o_5 belong to another.

The target relation network can be constructed as:

	O_1	O_2	O_3	O_4	O_5
O_1	*	1	1	0	0
O_2	1	*	1	0	0
O_3	1	1	*	0	0
O_4	0	0	0	*	1
O_5	0	0	0	1	*

The $*$ in the relation matrix means that we do not consider the self-relation strength. The four basic relation matrices (corresponding to these objects) in the social networks are shown in Fig. 1.

We can find that $0M_1+10M_2+10M_3+10M_4$ can exactly match the example relation matrix. However, such extracted relation might not a good approximation to the hidden relation on the whole object set. The relation M1 is more likely to be what the user desires. This is exactly the problem of unconstrained linear regression. We need to use some coefficient shrinkage techniques to solve such problem[17].

Thus, for each relation network, we normalize all the weights on the edges in the range [0, 1].

And, we put a constraint: $\sum_{i=1}^{n} a_i^2 \le 1$ on the objective function (2). Finally, our algorithm tries to solve the following minimization problem,

$$a^{opt} = \arg\min_{a} \left\| \tilde{v} - \sum_{i=1}^{n} a_i e_i \right\| \quad (5)$$

Subject to $\sum_{i=1}^{n} a_i^2 \le 1$

Such a constrained regression is called Ridge Regression[15] and can be solved by some numerical methods[4]. When we use such constrained relation extraction, the coefficients of the extracted relation for the above example are 1, 0, 0, 0. This shows that our constrained relation extraction can really solve the problem.

A MinCut-based algorithm: In the last subsection, we have presented a general method for exacting the hidden relation based on regression model. However, this method may fail when the examples provided by the user belong to only one community, which is referred to single community issue in the rest of this research. We provide an intuitive example in the following.

Suppose we have a user query (o_1, o_2, o_3, o_4, o_5), which belong to the same community. In the following two relations shown in Fig. 2a and b, regression model would prefer the relation M_1, since the higher connectivity between o_1, o_2, o_3, o_4 achieves a lower square error to the target relation. However, in relation M_1, the connectivity between o_5 and the other four examples are very weak. As can be seen, the connectivity in M_2 is much more uniform than that in

	O_1	O_2	O_3	O_4	O_5
O_1	*	0.8	0.7	0	0
O_2	0.8	*	0.9	0	0
O_3	0.7	0.9	*	0	0
O_4	0	0	0	*	0.6
O_5	0	0	0	0.6	*

(a) Relation M_1

	O_1	O_2	O_3	O_4	O_5
O_1	*	0	0.1	0	0
O_2	0	*	0	0	0
O_3	0.1	0	*	0	0
O_4	0	0	0	*	0
O_5	0	0	0	0	*

(b) Relation M_2

	O_1	O_2	O_3	O_4	O_5
O_1	*	0.1	0	0	0
O_2	0.1	*	0	0	0
O_3	0.1	0	*	0	0
O_4	0	0	0	*	0.1
O_5	0	0	0	0.1	*

(c) Relation M_3

	O_1	O_2	O_3	O_4	O_5
O_1	*	0	0	0	0
O_2	0	*	0.1	0	0
O_3	0	0.1	*	0	0
O_4	0	0	0	*	0.1
O_5	0	0	0	0.1	*

(d) Relation M_4

Fig. 1: The four basic relation matrices corresponding to the examples

	O_1	O_2	O_3	O_4	O_5
O_1	*	0	0	0	0
O_2	0	*	0.1	0	0
O_3	0	0.1	*	0	0
O_4	0	0	0	*	0.1
O_5	0	0	0	0.1	*

(a) Relation M_1

	O_1	O_2	O_3	O_4	O_5
O_1	*	0.4	0.4	0.5	0.5
O_2	0.4	*	0.3	0.2	0.4
O_3	0.4	0.3	*	0.3	0.3
O_4	0.5	0.2	0.3	*	0.4
O_5	0.5	0.4	0.3	0.4	*

(b) Relation M_2

Fig. 2: Two existing relations

M_1 while it has comparable strength. Therefore, M_2 should be a better choice for this user query. Unfortunately, the square error of M_2 is larger than that of M_1. This shows that the regression model may fail in such a case.

In order to deal with the single community issue, we need to take into account the weakest connection in the extracted relation. By graph theory, the value of the minimum cut on the graph can be used to evaluate the tightness of the graph.

Let G denote a weighted graph with weight matrix M. Let m denote the number of vertices.

A cut on the graph is defined as a set of edges which separates the vertices into two disconnected groups denoted by A and B such that $A \cap B = \varnothing$ and $A \cup B = D$. Thus, the value of the cut is: $\mathrm{cut}(G) = \sum_{i \in A} \sum_{j \in B} M(i, j)$

It is easy to see that there are totally 2^m-2 different cuts. Let cutk(G) = (A_k, B_k) denote the k-th cut. The minimum cut is defined as: $\min \mathrm{cut}(G) = \min_k \{\mathrm{cut}_k(G)\}$

If a graph can be easily cut into two subgraphs, it has a small minimum cut value. As an extreme case, the minimum cut value of a disconnected graph is 0. Naturally, the optimal extracted relation graph should have a large minimum cut value. Thus, for single community issue, we try to extract the optimal relation graph by maximizing its minimum cut value.

Let G_i, i = 1,…,n, denote the existing relation graphs defined only on the user query examples and Mi denote the corresponding weight matrices. Let a = $[a_1, a_2, \ldots, a_n]^T \in R_n$ denote the combination coefficients for different graphs. Thus $M = \sum_{i=1}^{n} a_i M_i$ is the weight matrix of the combined relation graph G. Let mincut(G) denote the minimum cut value of G. Our objective function can be written as follows:

$$a^{opt} = \arg\min_a \left\| \tilde{M} - \sum_{i=1}^{n} a_i M_i \right\|^2 \quad (6)$$

Generally, the minimum cut problem is an NP-hard problem. Thus the optimization problem (6) cannot be easily solved. However, in our problems, the number of examples provided by the user is usually small. That is, m is small, typically less than 10. Thus we can use linear programming techniques to solve the optimization problem (6) by the following derivation:
mincut(G)

$$= \min_{1 \le k \le 2^m - 2} \{\mathrm{cut}_k(G)\}$$

$$= \min_{1 \le k \le 2^m - 2} \left\{ \sum_{i \in A(k)} \sum_{j \in B(k)} \left(\sum_{l=1} a_l M_l(i, j) \right) \right\}$$

$$= \min_{1 \le k \le 2^m - 2} \left\{ \sum_{i \in A(k)} a_l \left(\sum_{i \in A(k)} \sum_{j \in B(k)} M_l(i, j) \right) \right\}$$

$$= \min_{1 \le k \le 2^m - 2} \left\{ \sum_{i \in A(k)} a_l . \mathrm{cut}_k(G_l) \right\}$$

Let v = mincut(G). The optimization problem in Eq. 6 can be reduced to the following linear programming problem:
Max v

$$\sum_{l=1}^{n} a_l . \mathrm{cut}_k(G_l) - v \ge 0, \quad (1 \le k \le 2^m - 2) \ (*)$$

$$\sum_{l=1}^{n} a_l = 1$$

$$a_l \ge 0, \ (1 \le l \le n)$$

With the constraints (*), v is guaranteed to be the minimum cut value and by maximizing v we can obtain the optimal combination coefficients ai. The number of constraints in this problem is 2^m-2+n+1, where m is the number of user-provided examples which is usually less than 10 and n is the number of existing relations. The above problem can be efficiently solved by linear programming techniques[3].

The proposed regression based algorithm and the MinCut based algorithm are used under different situations. When a user provides multiple community examples, regression-based algorithm can be used to find the best combination; when he provides single

community examples, MinCut-based algorithm can be used.

DISCUSSION

Since mining hidden communities in heterogeneous networks represents a promising research direction, there are many issues that need to be discussed. Here we focus on the problem solving philosophy.

First, one may wonder the complexity at comprehension and combination of multiple social networks in the analysis. We do agree that multiple social networks form complex, multiple, interrelated graphs and with the massive amount of data mounting, it is challenging for anyone to grasp the whole picture of such dynamic, evolving social networks and work out a balanced combination of multiple networks for a particular user query. However, such multiple networks do exist and it is inappropriate to blindly merge them into one since different networks plays different roles in particular queries, as shown in our experiments. Therefore, we believe that developing new multi-network mining algorithms to dynamically combine multiple relevant networks to form combined virtual networks based on user's example queries is a new and appropriate problem solving methodology.

Second, since it is difficult for a user to comprehend the whole picture of numerous social networks, one may wonder how a user is able to pose high-quality queries. Based on our experience, although it is difficult for a user to comprehend the overall multiple networks, a user usually has good knowledge on a small set of examples (such as influential researchers, movie/sport stars, big companies, or popular commodities). Such firm grasp of a small set of examples is often sufficient to pose intelligent queries, learn additional facts and form informative combined networks.

Third, one may wonder how to comprehend the answers returned from such a network analysis.

Since a derived hidden network is a weighted matrix as a combination of multiple existing networks, it is often difficult to understand the minor weight differences in the results. However, the real essence is at the new facts derived from such hidden networks and their associated rankings. This resembles Google-like keyword-based Web search. It is not so crucial to understand the derived Web linkage weighting and claim it is optimal. However, the return of quality rankings on the interesting results demonstrate its utility.

CONCLUSION

Different from most social network analysis studies, we assume that there exist multiple, heterogeneous social networks and the sophisticated combinations of such heterogeneous social networks may generate important new relationships that may better fit user's information need. Therefore, our approach to social network analysis and community mining represents a major shift in methodology from the traditional one, a shift from single-network, user-independent analysis to multi-network, user-dependant and query-based analysis. Our argument for such a shift is clear: multiple, heterogeneous social networks are ubiquitous in the real world and they usually jointly affect people's social activities.

Based on such a philosophy, we worked out a new methodology for relation extraction and proposed two algorithms in different situations. With such query-dependent relation extraction and community mining, fine and subtle semantics are captured effectively. Our discussion also shows it is expected that the query-based relation extraction and community mining would give rise to a lot of potential new applications in social network analysis.

There are a lot of issues that need to be studied further. First, our approach adopts a regression-based graph matrix analysis approach. There are potentially many other approaches that can be explored and compared with this approach. We will expect that future studies may propose even more powerful approaches in relation extraction than what is proposed here.

Second, our relation extraction algorithm has made a lot of simplifications in the analysis. In general, links within the same network and among different networks may carry different weights. For example, one can imagine that the links among co-author networks should be inherently stronger than those among co-proceedings since average size (# of links) in the co-author group is much smaller than that in the co-proceedings group. This is not considered in our simple model. Thus we expect the prediction power will be substantially enhanced if such information is incorporated in the new algorithm.

Third, our query model considers only one simple group of nodes (such as researchers). A more powerful query model may involve and, or, not operators on those groups. For example, one may like to find those who co-attend the same conference but never co-authored a paper using the not operator. This will be useful for finding referees for conference submissions. These issues may form an exciting frontier for future research.

REFERENCES

1. Adamic, L.A. and E. Adar, 2002. Friends and neighbors on the web. Technical Report, Xerox Parc.
2. Agrawal, R., S. Rajagopalan, R. Srikant and Y. Xu, 2003. Mining newsgroups using networks arising from social behavior. In: Proceedings of 12th International World Wide Web Conference.
3. Bazaraa, M., J. Jarvis and H. Sherali, 2004. Linear Programming and Network Flows. Wiley, 3rd Edn.
4. Bjorck, A., 1996. Numerical Methods for Least Squares Problems. SIAM.
5. Cai, D., X. He, J.R. Wen and W.Y. Ma, 2004. Block-level link analysis. In: Proceedings of the ACM SIGIR'2004.
6. Chakrabarti, S., 2001. Integrating the document object model with hyperlinks for enhanced topic distillation and information extraction. In: Proceedings of the 10th International World Wide Web Conference.
7. Chakrabarti, S., M. van den Berg and B. Dom, 1999. Focused crawling: A new approach to topic-specific web resource discovery. In: Proceedings of The 8th International World Wide Web Conference.
8. Chen, C. and L. Carr, 1999. Trailblazing the literature of hypertext: Author co-citation analysis (1989-1998). In: Proceedings of the 10th ACM Conference on Hypertext and hypermedia.
9. Cohn, D. and H. Chang, 2000. Learning to probabilistically identify authoritative documents. In: Proceedings of the 17th International Conference on Machine Learning.
10. Cook, D.J. and L.B. Holder, 2000. Graph-based data mining. IEEE Intel. Syst., 15 (2): 32-41.
11. Domingos, P. and M. Richardson, 2001. Mining the network value of customers. In: Proceedings of the 7th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 57-66. ACM Press.
12. Flake, G.W., S. Lawrence and C.L. Giles, 2000. Efficient identification of web communities. In: Proceedings of the 6th International Conference on Knowledge Discovery and Data Mining (ACM SIGKDD-2000).
13. Flake, G.W., S. Lawrence, C.L. Giles and F. Coetzee, 2002. Self-organization of the web and identification of communities. IEEE Comput., 35 (3): 66-71.
14. Gibson, D., J. Kleinberg and P. Raghavan, 1998. Inferring web communities from link topology. In: Proceedings of the 9th ACM Conference on Hypertext and Hypermedia.
15. Hastie, T., R. Tibshirani and J.H. Friedman, 2001. The Elements of Statistical Learning. Springer-Verlag.
16. Haveliwala, T., 2002. Topic-sensitive pagerank. In: Proceedings of the 11th International World Wide Web Conference.
17. He, X. and P. Niyogi, 2003. Locality preserving projections. Adv. Neural Inform. Processing Syst., 16.
18. Kautz, H., B. Selman and A. Milewski, 1996. Agent amplified communication. In: Proceedings of AAAI-96, pp: 3-9.
19. Kautz, H., B. Selman and M. Shah, 1997. Referral web: Combining social networks and collaborative filtering. Commun. ACM., 40 (3): 63-65.
20. Kleinberg, J., 1999. Authoritative sources in a hyperlinked environment. J. ACM., 46 (5): 604-622.
21. Kleinberg, J.M., 1999. Hubs, authorities and communities. ACM Comput. Surveys, 31 (4).
22. Kumar, R., P. Raghavan, S. Rajagopalan and A. Tomkins, 1999. Trawling the web for emerging cyber communities. In: Proceedings of The 8th International World Wide Web Conference.
23. Milgram, S., 1967. The small world problem. Psychol. Today, 2: 60-67.
24. Page, L., S. Brin, R. Motwani and T. Winograd, 1998. The pagerank citation ranking: Bringing order to the web. Technical Report, Stanford University.
25. Popescul, A., G.W. Flake, S. Lawrence, L.H. Ungar and C.L. Giles, 2000. Clustering and identifying temporal trends in document databases. In IEEE Advances in Digital Libraries.
26. Schwartz, M.F. and D.C.M. Wood, 1993. Discovering shared interests using graph analysis. Commun. ACM., 36 (8): 78-89.
27. Toyoda, M. and M. Kitsuregawa, 2002. Observing evolution of web communities. In: Proceedings of the 11th International World Wide Web Conference (WWW2002).
28. Washio, T. and H. Motoda, 2003. State of the art of graph-based data mining. SIGKDD Explor. Newsl., 5(1): 59-68.
29. Wasserman, S. and K. Faust, 1994. Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge, UK.
30. Yin, X., J. Han, J. Yang and P.S. Yu, 2004. Crossmine: Efficient classification across multiple database relations. In: Proc. 2004 Int. Conf. on Data Engineering (ICDE'04).
31. Zhou, W.J., J.R. Wen, W.Y. Ma and H.J. Zhang, 2002. A concentric-circle model for community mining. Technical Report, Microsoft Research.