Journal of Theoretical and Applied Information Technology
28th February 2013. Vol. 48 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MINING TRIBE BASED ON THE FREQUENT PATTERN

[1]LINA LIU, [1]JIHONG SHEN, [2]QIANGHUA ZHU

[1]College of automation, Harbin Engineering University, Harbin 150001, Heilongjiang, China
[2]College of electronic engineering, Naval University of Engineering, Wuhan 430033, Hubei, China

ABSTRACT

Discovering the influencers in the new social network is very important for the promotion of products and the supervision of public opinion. Most of the previous research was based on the method of mining influential individuals, while the tribe-leaders were neglected. In this paper, a new method of mining tribe-leaders is proposed based on the frequent pattern of propagation. First, a method of changing the diffusion trees is proposed to overcome the problem of multi-pattern in propagation, where the information propagation trees are changed into a connected undirected acyclic graph. Then, a new frequent subgraph mining method called Tribe-FGM is proposed to improve the efficiency of graph mining by reducing the scale of pattern growth. Experiments are conducted on a real dataset, and the results show that Tribe-FGM is more effective than the method of Unot.

Keywords: *Social Network, Frequent pattern, Microblog, Graph Mining*

1. INTRODUCTION

Recently, online social networks have obtained considerable popularity and are now among the most popular sites on the Web, such as blog, forum, microblog, etc. Online social networks are organized by many users and links between users who are friends or acquaintances. Users join a network, publish their profiles and any other content, and create links to other users with whom they associate. Social networks provide a basic environment for maintaining social relationships, finding users with similar interests, and posting content or comments contributed or endorsed by other users. In addition, social networks play an important role for the spread of information since a piece of information can propagate from one node to another through a link on the network in the form of "word-of-mouth" communication. Therefore, Social network sites have become one of the several main sites where people spend most of their time [1]. Their massive popularity has led to the viral marketing of content, products, or political campaigns on the sites. For instance, if we know there are a small number of "leaders" who set the trend for various actions, targeting them for the adoption of new products or technology could be profit-able to the companies. Also, if we know there are some "opinion leaderships" who set the trend for public opinions, we can target them to control the propagation of the public opinion. So, it is important to discover influential individuals for the promotion of products and the supervision of public opinions.

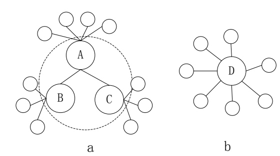

Figure 1: The Tribe-Leaders And The Influential Individuals

A number of recent empirical papers have addressed the matter of diffusion on networks in general, and particularly the attributes and roles of influencers. In general, influencers are loosely defined as individuals who disproportionately impact the spread of information. Interest among researchers and marketers alike has increasingly focused on whether or not diffusion can be maximized by seeding a piece of information or a new product with certain special individuals, often called "influentials" or simply "influencers". Also, other research has mainly focused on discovering influential individuals such as leaderships [2] and ranking web documents [3, 4] and users [5].

Most of the previous research was based on the mining of influential individuals, while the tribe-leaders were ignored neglected. A Tribe-leader is a set of influential individuals who exchange information actively and frequently. Clearly, every tribe leader is an influential individual and all tribe leaders form a connected undirected acyclic graph. Tribe-leaders play a very important role in setting the trend for advertisement and public opinions. If an influential individual in the tribe-leader is affected, it will immediately diffuse the information to other influential individuals. This means the tribe-leader will diffuse the information to other

Journal of Theoretical and Applied Information Technology
28th February 2013. Vol. 48 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

nodes in the social network through all the influential individuals in the tribe-leader. Obviously, the tribe-leaders have better ability to diffuse information than a single influential individual. A single influential individual in the tribe-leader may not necessarily have a strong ability to diffuse information as other influential individuals, but the whole ability of a tribe-leader is stronger than that of other influential individuals. For example, the tribe-leader {A, B, C} in figure 1(a) has better ability to diffuse information than the influential individual D in figure 1(b). Therefore, mining the tribe-leaders is more important than discovering the influential individuals.

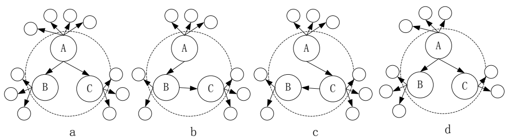

Figure 2: The Four Diffusion Trees About A Topic In The Microblog

To discover the tribe-leader, we use the frequent pattern of propagation which was neglected in previous research. The propagation of information in the social network constructs diffusion trees, and the influential individuals in some diffusion trees diffuse the information frequently through their interactions. For instance, figure 2 shows four diffusion trees about a topic in the microblog. In the four diffusion trees, influential individuals A, B, C diffuse information frequently through the interactions of themselves. So, the set of {A, B, C} is a tribe-leader.

In this paper, we use the frequent pattern of propagation in the online social net-work. First, a method of changing the diffusion trees is proposed to overcome the problem of multi-pattern in propagation, where the information propagation trees are changed into a connected undirected acyclic graph. Then considering its support and strength, a new frequent subgraph mining method called Tribe-FGM is proposed to improve the efficiency of graph mining by reducing the scale of pattern growth. Finally, we validate the effectiveness of our method by comparing it with the repost algorithms in the real dataset of sina microblog from china. Experimental results indicate that the tribe-leaders with our method are consistently better than that of repost algorithms in both the one-step and multi-step coverage.

2. PROBLEM DEFINITION

A social graph is an undirected graph G = (V, E) where the nodes are users. There is an undirected edge between users u and v representing a social tie between the users. The tie may be an explicit repost relationship in the microblog.

The propagation of information in the social network will construct diffusion trees $T=(V,E,\Sigma,L,r)$ where the nodes are users. There is a directed edge between users u and v representing a diffused path between the users. The alphabet Σ is a set of labels, and the mapping label: $L:V\cup E\rightarrow\Sigma$ is called a labeling function. We define the label of edges as the name of information diffused. For instance, the propagation of information is a post in the microblog. The alphabet r represents the root node of diffusion trees. For every node v ∈ V, there is a unique path UP (v) = ($v0 = r, v_1. . . v_d$) ($d \geq 0$) from the root r to v. Let u and v be nodes. If (u, v) ∈ E then u is a parent of v, or v is a child of u. If there is a path from u to v, then u is an ancestor of v, or v is a descendant of u. A leaf is a node having no child.

Property 1: diffusion trees $T=(V,E,\Sigma,L,r)$ are rooted unordered trees, where there are no nodes with same labels in a diffusion tree.

Property 2: diffusion trees $T=(V,E,\Sigma,L,r)$ are connected directed acyclic graph.

The propagation of multi-information in the social network will construct multi-diffusion trees called diffusion forest $F=\{T_1,T_2,...,T_n\}$. For example, the propagation of multi-post about a topic in the microblog will construct a diffusion forest.

In general, influencers are loosely defined as individuals who disproportionately impact the spread of information. Unfortunately, this definition is fraught with ambiguity regarding the nature of the influence in question, and hence the type of individuals who might be considered special. In light of this definitional ambiguity, we note, however, that our use of the term influencer corresponds to a particular and somewhat narrow definition of influence, specifically the user's ability to write the post which diffuses directly through the social network graph.

Definition 1 (influence). Give a diffusion tree $T=(V,E,\Sigma,L,r)$; the influence of a node u in the diffusion tree is defined as $\inf luence_u=|L(u)|$, where $L(u)$ is a set whose elements are those nodes which are linked with node u directly.

In the diffusion forest $F=\{T_1,T_2,...,T_n\}$, a user may write lots of posts which diffuse directly through the social network graph in multi-diffusion trees such as node A in figure 2. So, influential individuals are defined as those users who have high influence and appear in multi-diffusion trees.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

For example, influential individuals about a topic in microblog are those users who write lots of posts about a topic and most of the posts have high influence. So, two thresholds σ and ψ are assigned, where the threshold σ represents the frequency of a user appearing in the multi-diffusion trees and the threshold ψ represents the influence of the user. According to property 1, we can infer that a user appears only once in a diffusion tree. So, the frequency of a user appearing in the multi-diffusion trees is the number of diffusion trees which contain this user.

Definition 2 (influential individuals). Give a social network graph G = (V, E) and the diffusion forest $F = \{T_1, T_2, ..., T_n\}$ from the social network graph, and two thresholds σ and ψ, a user v ∈ V is an influential individual iff:

$$|L(T_i)| \geq \sigma, v \in T_i \ and \ \inf luence_{v,T_i} \geq \psi, v \in T_i \qquad (1)$$

The set $L(T_i)$ represents the set of multi-diffusion trees containing the user v. The formula $\inf luence_{v,T_i}$ represents the influence of the user v in the diffusion tree T_i .

A Tribe-leader is a set of influential individuals who exchange information actively and frequently. Clearly, every leader is an influential individual and a tribe-leader is a connected undirected acyclic graph. So, three thresholds σ , ψ and ζ are assigned, where the threshold σ represents the frequency of a tribe-leader appearing in the multi-diffusion trees, the threshold ψ represents the influence of the user in this tribe-leader, and the threshold ζ represents the whole influence of all users in this tribe-leader.

Definition 3 (tribe-leader). Give a social network graph G = (V, E) and the diffusion forest $F = \{T_1, T_2, ..., T_n\}$ from the social network graph, and three thresholds σ , ψ and ζ , a set of user $L(v_i)$ is a tribe-leader iff:

$$\exists G_{tribe} = (V_{tribe}, E_{tribe}) \ (V_{tribe} = L(v_i) \text{ and } \forall v_i, v_j \in V_{tribe}, \exists (v_i, v_j) \in E_{tribe})$$
$$and \ |L(T_i)| \geq \sigma, L(v_i) \subseteq T_i \ and \ \inf luence_{v,T_i} \geq \psi, v \in T_i, v \in L(v_i)$$
$$and \sum_{v_k \in L(v_i)} \inf luence_{v_k,T_i} \geq \zeta, v_k \in T_i \qquad (2)$$

Give a social network graph G = (V, E) and multi-diffusion trees from the social network graph, the discovery of tribe-leaders is formalized as a problem of frequent pattern mining. Intuitively, we can discover tribe-leaders through the method of frequent subtree mining. However, in fact, a tribe-leader exist multi-pattern of propagation. For instance, there are three patterns of propagation in the users' set {A, B, C} in figure 3. The three users A, B and C exchange information actively and frequently, and also three of these users are influential individuals. According the definition 3, we can infer the users' set {A, B, C} is a tribe-leader. The methods of frequent subtree mining only consider single paths of propagation, but the multi-pattern of propagation in the social network is neglected. So, some of members in the tribe-leader will be neglected for specific support. For instance, given the support thresholds 50%, in the figure 3, we can infer the set {A, B} and the set {B, C} are two tribe-leaders according the method of frequent subtree mining not considering the multi-pattern of propagation in the social network. More, if we give the support thresholds 100%, the tribe-leader {A, B, C} will be neglected. To overcome the problem of the multi-pattern of propagation, we change the diffusion trees into connected undirected acyclic graphs, which contain two steps: eliminating the direction of the diffusion trees and adding some edges which exist in the social network and also whose nodes are affected in the spread of information into the diffusion graphs.

For instance, we can get connected undirected acyclic graphs called diffusion graphs b, c and d in figure 4 after eliminating the direction of the diffusion trees. The two diffusion paths $A \rightarrow B$ and $B \rightarrow A$ are eliminated by eliminating the direction of the diffusion trees. Given the support thresholds 100%, we can infer the set {A, B} is a tribe-leader, where the results are better than that of the diffusion trees without eliminating the direction.

Next, we can add some edges which exist in the social network and also whose nodes are affected in the spread of information into the diffusion graphs. For instance, in figure 4(b), node B and node C are affected by node A; moreover, there is an edge between node B and node C in the social network in figure 4(a). So, we can add an edge into the diffusion graph in figure 4(b) and then get the changed diffusion graph in figure 5(b). With the same method, after adding the edge into the diffusion graph in figure 4(c) and figure 4(d), we can get the changed diffusion graphs in figure 5(c) and figure 5(d). After adding the edges into the diffusion graphs in figure 5, we can find that the multi-diffusion paths are eliminated. Thus, given the support thresholds 100%, we can infer the set {A, B, C} is a tribe-leader, where the results are better than that of the diffusion graphs without adding some edges into the diffusion graphs.

Journal of Theoretical and Applied Information Technology
28th February 2013. Vol. 48 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

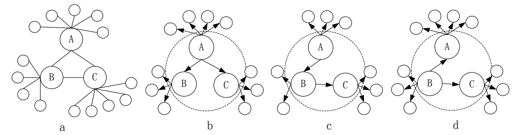

Figure 3: The Three Patterns Of Propagation In The User Set {A, B, C}

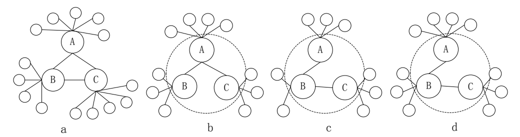

Figure 4: The Diffusion Graphs After Eliminating The Direction Of The Diffusion Trees

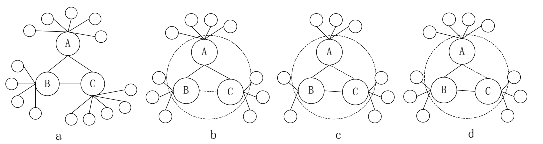

Figure 5: The Changed Diffusion Graphs After Adding Some Edges Into The Diffusion Graphs

3. THE METHODS OF CHANGING DIFFUSION TREES

3.1 Eliminating The Direction Of Diffusion Trees

To overcome the problem of multi-pattern in propagation, we first eliminate the direction of diffusion trees and change the diffusion trees into connected undirected acyclic graphs. To eliminate the direction of diffusion trees, the diffusion trees are formalized as adjacency matrixes.

$$A[u,v]=\begin{cases}1 & \exists edge(\mathrm{u},\mathrm{v})\\ 0 & else\end{cases} \qquad (3)$$

According to the calculation of matrixes, we can eliminate the direction of diffusion trees. We define the diffusion graphs after eliminating the direction of diffusion trees as $GT=(V,E,\sum,L)$. Also, the diffusion graphs are formalized as adjacency matrixes Q, that is $Q=A+A^T$.

Algorithm 1. The method of adding some edges into diffusion graphs

Input.

1) The social network graph G = (V, E), adjacency matrixes M, and the sequence table S.

2) The set of diffusion graphs $\{GT_1,GT_2,...,GT_n\}$, adjacency matrixes Q_i, and the sequence table ST_i of each diffusion graphs.

Output.

The set of changed diffusion graphs $\{GF_1,GF_2,...,GF_n\}$, adjacency matrixes N_i, and the sequence table SF_i of each changed diffusion graphs.

Method.

1) For each graph GT_i in the set of diffusion graphs $\{GT_1,GT_2,...,GT_n\}$.

2) Getting the set of nodes' labels $L(name_i)$ from the sequence table ST_i.

3) Getting the set of sequence $L(list_i)$ from the sequence table S according to the set of nodes' labels the $L(name_i)$.

4) Getting the relationship of edges of changed diffusion graphs from the adjacency matrixes M according to the set of sequence $L(list_i)$.

5) Storing the sequence table SF_i with the set of nodes' labels $L(name_i)$ and storing the adjacency matrixes N_i with the relationship of edges getting from 4).

6) End for

7) Return

Property 3: the diffusion graphs $GT=(V,E,\sum,L)$ are connected undirected acyclic graphs, and are subgraphs of the social network graph G = (V, E).

3.2 Adding Some Edges Into Diffusion Graphs

To eliminate multi-diffusion paths in the diffusion graphs, we add into the diffusion graphs some edges which exist in the social network and also whose nodes are affected in the spread of information.

According to section 4.1, the diffusion graphs are defined as $GT=(V,E,\sum,L)$ and the diffusion graphs are formalized as adjacency matrixes Q. Also, we define a sequence table ST to store the labels of nodes. At the same time, we define the social network graph G = (V, E), adjacency matrixes M, and the sequence table S. And we define the changed diffusion graphs $GF=(V,E,\sum,L)$, adjacency matrixes N, and the sequence table SF. The method of adding some edges into diffusion graphs is proposed in table 1.

For each diffusion graph, we will add some edges into diffusion graphs and the time complexity of adjacency matrixes is $O(n\wedge 2)$. In this way, the time complexity of algorithm 1 is $O(n\wedge 3)$.

Lemma 1. The changed diffusion graphs $GF=(V,E,\sum,L)$ are subgraphs of the social network graph G = (V, E).

Proof. According to property 3, the diffusion graphs $GT=(V,E,\sum,L)$ are subgraphs of the social network graph G = (V, E). The set of nodes $L(gtv_i)$ in the diffusion graphs is the subset of the set of nodes $L(gv_i)$ in the social network, that is, $L(gtv_i)\subset L(gv_i)$. Similarly, the set of edges $L(gte_i)$ in diffusion graphs is the subset of the set of edges $L(ge_i)$ in the social network, that

Journal of Theoretical and Applied Information Technology
28th February 2013. Vol. 48 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

is, $L(gte_i) \subset L(ge_i)$. Also, all of the nodes which are linked by the set of edges $L(gte_i)$ are in the set of nodes $L(gtv_i)$. The set of nodes in the changed diffusion graphs $GF = (V, E, \Sigma, L)$ is the same as the set of nodes in the diffusion graphs. So, the set of nodes $L(gfv_i)$ in the changed diffusion graphs is the subset of the set of nodes $L(gv_i)$ in the social network, that is, $L(gfv_i) \subset L(gv_i)$. According to step 4 of algorithm 1, the set of edges $L(gfe_i)$ in changed diffusion graphs is a subset of $L(ge_i)$ in the social network, that is, $L(gfe_i) \subset L(ge_i)$. Also, all of the nodes which are linked by the set of edges $L(gfe_i)$ are in the set of nodes $L(gfv_i)$. So, the changed diffusion graphs $GF = (V, E, \Sigma, L)$ are the subgraphs of the social network graph G = (V, E).

Through changing the diffusion trees into connected undirected acyclic graphs, the set of diffusion trees is changed into the set of changing diffusion graphs. According to lemma 1, the changing diffusion graphs are the subgraphs of the social network graph. So, the discovery of tribe-leaders is formalized as a problem of frequent subgraph mining.

4. TRIBE-FGM

To discover the tribe-leaders, a new algorithm called Tribe-FGM is proposed in this paper. The algorithm of Tribe-FGM is a new method of frequent subgraph mining, which uses the support and the strength to improve the efficiency of graph mining by reducing the scale of pattern growth.

Algorithm 2. The algorithm of Tribe-FGM to discover the tribe-leaders

Input.

1) The set of changed diffusion graphs D, the influence of each user $\inf luence_{ij}$.

2) The thresholds of support $\min\_\sup$, strength $\min\_str$, ψ and ζ , and the DFS code s.

Output. The set of tribe-leaders S.

Method.

1) $S \leftarrow \phi$.

2) Calling the function of memberPruning ($\min\_str$, ψ , $\inf luence_{ij}$, D).

3) Calling the function of tP_gS ($\min\_\sup$, $\min\_str$, ψ , ζ , $\inf luence_{ij}$, D, s).

4) **procedure** memberPruning($\min\_str$, ψ , $\inf luence_{ij}$, D)

a) Computing the frequency f of users whose $\inf luence_{ij} \ge \psi$.

b) For each changed diffusion graph GF in the set of D, do

c) for each member in the changed diffusion graph, do

d) if $f / |D| < \min\_str$ for a user

e) Remove this user

f) End for

g) End for

h) Return the new set of changed diffusion graphs D.

5) **procedure** tP_gS($\min\_\sup$, $\min\_str$, ψ , ζ , $\inf luence_{ij}$,D,s)

a) Inserting the s into the S, and $C \leftarrow \phi$.

b) Finding all the edges e which are most right expanded $s\Diamond_r e$, after scanning the D.

c) Inserting the $s\Diamond_r e$ into the C.

d) Sorting the C according to the DFS.

e) For each $s\Diamond_r e$ in the C, do

f) Computing the frequency f1 of users in the $s\Diamond_r e$ whose $\sum_{v \in s\Diamond_r e} \inf_v(Q) \ge \zeta$

g) Computing the frequency f2 of users in the $s\Diamond_r e$

h) if $f1 / f2 \ge \min\_str \ and \ f2 / |D| \ge \min\_\sup$

i) tP_gS($\min\_\sup$, $\min\_str$, ψ , ζ , $\inf luence_{ij}$,D, $s\Diamond_r e$)

j) End for

k) Return

6) End

Definition 4 (support). Given the set of changed diffusion graphs $Graph = \{GF_1, GF_2, ..., GF_n\}$ and a tribe-leader T, the support $Supp_{Graph}(T)$ of the tribe-leader T in the set of changed diffusion graphs is defined as follows.

$$Supp_{Graph}(T) = \frac{|\{Q(c_i) \mid T \subseteq Q(c_i) \,\&\, Q(c_i) \in Graph\}|}{|Graph|} \tag{4}$$

The $|Graph|$ represents the size of the set of changed diffusion graphs. The range of the support is from 0 to 1. The support represents the frequency of the tribe-leaders appearing in the set of changed diffusion graphs. It also represents the activity of users in the tribe-leaders.

To measure the influence of tribe-leaders, we give the definition of strength. Given a specified support, we can get a set of frequent subgraphs. The sum of users' influences in different subgraphs is different. For example, the sum of users' influences from A, B, C in figure 3(b) is 10, and the influences of individual users are 4, 3 and 3 respectively. Also, the sum of users' influences from A, B, C in figure 3(c) is 8, and the influences of individual users are 3, 2 and 3 respectively. The strength represents the frequency of the tribe-leaders whose whole influence is higher than the specified thresholds.

Definition 5 (strength). Given the set of changed diffusion graphs $Graph = \{GF_1, GF_2, ..., GF_n\}$, a tribe-leader T, and the specified thresholds ψ and ζ , the strength $Stre_{Graph,\psi,\zeta}(T)$ of the tribe-leader T in the set of changed diffusion graphs is defined as.

$$Stre_{Graph,\psi,\zeta}(T) = \frac{|\{Q \mid \sum_{v \in Tribe} \inf luence_v(Q) \ge \zeta, \inf luence_v(Q) \ge \psi, T \subseteq Q, Q \in Graph\}|}{|\{Q \mid T \subseteq Q, Q \in Graph\}|} \tag{5}$$

Journal of Theoretical and Applied Information Technology
28th February 2013. Vol. 48 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The formula $\sum_{v \in Tribe} \inf_v(Q) \geq \zeta$ represents the whole influence of a tribe-leader, and the formula $\inf_v(Q) \geq \psi$ represents the influence of an individual user in the tribe-leader. The strength represents the frequency of the tribe-leaders whose whole influence is higher than the specified thresholds. The range of the strength is from 0 to 1. For example, given the threshold ζ as 10 and the threshold ψ as 2, the strength of the tribe-leader {A, B, C} in figure 5 is 66.7%. Given the threshold ζ as 8 and the threshold ψ as 2, the strength of the tribe-leader {A, B, C} in figure 5 is 100%.

Based on the algorithm of gSpan [6], we propose a new algorithm of frequent subgraph mining called Tribe-FGM which uses the support and the strength to improve the efficiency of graph mining by reducing the scale of pattern growth. Traditional methods of frequent subgraph mining such as gSpan did not consider the strength and would add some low influence nodes into the candidate thus increase the size of searching space. So, a new algorithm called Tribe-FGM is proposed in this paper.

Most of the nodes whose influence is low are pruned by the function of memberPruning (). The experimental results in section 6 indicate that only a small number of users are influential individuals, most of the users only participate in the discussion of topics, and many non-influential individuals form the "the long tail". So, most of the nodes are pruned by the function of memberPruning () and the scale of pattern growth will be reduced. The time complexity of the function of memberPruning () is $O(n \wedge 2)$. The support and the strength are considered in the function of tP_gS () to improve the efficiency of graph mining by reducing the scale of pattern growth.

5. EXPERIMENTS

We validate the performance of the algorithm of Tribe-FGM and the total number of influential individuals in the tribe-leaders with different parameters. We give the thresholds as follows: $\zeta = 8$, $\psi = 5$.

First, we present different experimental results with different supports, where the threshold of the strength $\min\_str = 80\%$.

In the experiments, the performance of the algorithm with a specified support is an average time-consuming of 10 experiments with the specified support.

As shown in figure 6, 7, the higher the support, the lower the time-consuming of the algorithm and the less the total number of influential individuals in the tribe-leaders. As the scale of pattern growth will decrease when the support increases, the time-consuming of the algorithm becomes lower and the total number of influential individuals becomes less.

Then, we present different experimental results with different strengths, where the threshold of the support $\min\_sup_{earthquake} = 0.7\%$, $\min\_sup_{two\ sessions} = 1\%$.

In the experiments, the performance of the algorithm with a specified strength is an average time-consuming of 10 experiments with the specified strength.

As can be seen from figure 8, 9, the higher the strength, the lower the time-consuming of the algorithm and the less the total number of influential individuals in the tribe-leaders. As more nodes will be pruned and the scale of pattern growth will decrease when the strength increases, the time-consuming of the algorithm becomes lower and the total number of influential individuals becomes less.

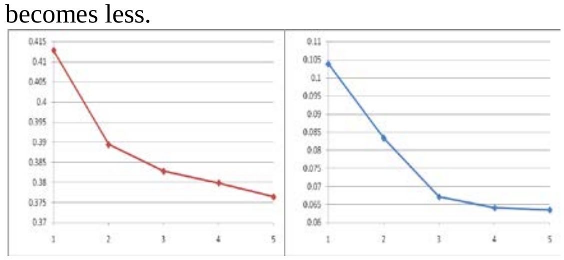

Figure 6: The Performance With Different Supports

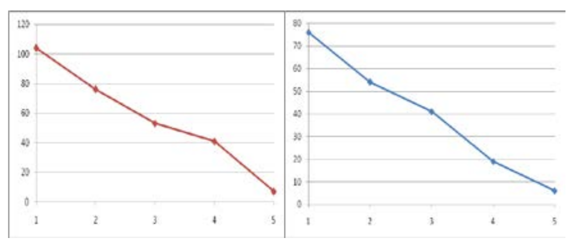

Figure 7: The Number Of Influential Individuals With Different Supports

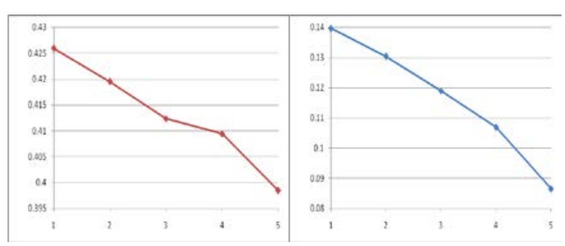

Figure 8: The Performance With Different Strengths

Journal of Theoretical and Applied Information Technology
28th February 2013. Vol. 48 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

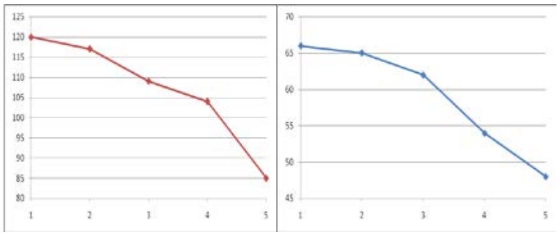

Figure 9: The Number Of Influential Individuals With Different Strengths

6. CONCLUSION

In this paper, a new method of mining tribe-leaders is proposed based on the frequent pattern of propagation. The algorithm of Tribe-FGM is a new method of frequent sub-graph mining, which uses the support and the strength to improve the efficiency of graph mining by reducing the scale of pattern growth.

ACKNOWLEDGEMENTS

This work was supported by NSFC (61202127)

REFERENCES:

[1] Y. Ahn, S. Han, H. Kwak, et al., "Analysis of topological characteristics of huge online social networking services", Proceedings of the 16th International World Wide Web Conference, New York: ACM, April 4-7, 2007, pp. 835-844.

[2] D. Wu, Y. ke, J. X. Yu et al., "Leadership discovery when data correlatively evolve", *World Wide Web Journal*, Vol. 14, No. 1, 2011, pp. 1-25.

[3] D. Kempe, J. Kleinberg, É. Tardos, "Maximizing the spread of influence through a social network", Proceedings of the 9th ACM SIGKDD international conference on Knowledge discovery and data mining, New York: ACM, May 15-18, 2003, pp. 137-146.

[4] E. Bakshy, J. -M. Hofman, W. -A. Mason, et al., "Everyone's an Influencer: Quantifying Influence on Twitter", Proceedings of the 4th ACM International Conference on Web Search and Data Mining, New York: ACM, June 11-13, 2011, pp. 65-74.

[5] J. Goldenberg, B. Libai, E. Muller, "Talk of the network: A complex systems look at the underlying process of word-of-mouth", *Marketing Letters*, Vol. 12, No. 3, 2001, pp. 211-223.

[6] X. Yan, J. Han, "gSpan : Graph-based Substructure Patterns Mining", Proceedings of IEEE the second International Conference on Data Mining, NJ: IEEE, July 8-10, 2002, pp. 721-724.