# Generalized Analysis of Message Propagation on Social Network

Min-Feng Wang, Chi-Sheng Huang*, Meng-Feng Tsai, Bo-Ru Song,
Shin-Fu Su and Cheng-Hsien Tang

*Department of Computer Science and Information Engineering, National Central University, No.300, Jhongda Rd., Jhongli City, Taoyuan County, Taiwan, ROC*
*{945402023, 100582022, mftsai, 985202107, 995202016, 965202060}@cc.ncu.edu.tw*

***Abstract***

*The importance of social network analysis has become more remarkable in recent years. The social network researches which focus on the behavior of users propagating messages have been growing in recent years. Besides, most of the researches of influence evaluation in the past focused on the behavior of propagating between users, instead of considering the behavior of propagating between different social communities. Hence, this research provides a system to evaluate the pattern of propagating messages. The information propagation, however, can take place at different levels of social groups-among individuals and among groups or communities. Adapting from a generalized suffix tree, our proposed methodology can be constructed to form a framework for querying various kinds of related propagation effects, and the effects can be observed corresponding to different levels of granularities of social communities.*

***Keywords:*** *Social Network Analysis, Suffix Tree, Information Propagation*

## 1. Introduction

Pervasiveness of social network services has led to the generation of new customer groups enabling enterprises to adopt new marketing strategy to develop, market and support of products and services in the Internet. Well-known social websites, including Facebook, Twitter and Digg and others etc., allow users to construct a personal profile, share interesting information with other people, and build relationships within a community. The mode of interaction on social websites is rapidly affecting people's social behaviors and consumer habits. These social websites are viewed as a platform for business to spread advertisements and other promotional material. According to eMarketer's forecasts [1], US marketers will spend 3.93 billion to advertise on social networking sites in 2012. Spending will be up 27.7% over the 3.08 billion advertisers devoted to social networks in 2011. Therefore, the last few years have seen considerable work on social network analysis. Especially in the social marketing discipline, much research tends to apply statistics, modeling, and analytics to measure and investigate how the information can be propagated over a social network efficiently and widely [2, 3]. However, information diffusion encompasses everything from product quotations and contracts to customer service and marketing advertisements. Many businesses arbitrarily spread massive unnecessary messages to all people through social websites without considering users' interests and requirements. This situation will not only increase advertising cost of social marketing, but also will reduce the focus on the target customer on social networks. Furthermore, lack of research in the past has focused on designing an efficient and flexible method for answering and mining the questions in social networks to support social marketing and target marketing, such as "Who are all influenced by a specific user or community in a limited time?", or even "Who are the top-K users or communities with the most number of distinct paths can propagate message?" Although many marketing techniques may be used to spread information over a social network, the target consumers should be defined, and the relative suitable messages should be disseminated to them in a certain time period. Consequently, enterprises need a tool to analyze message propagation behavior at different combinations of community and time dimensions.

Therefore, this work adopts the community detection method to generate a community-based concept hierarchy [4], and it is integrated into the developed generalized suffix tree data structure for mining and querying nodes and paths in a social network. In addition, we also apply the concept of web session to set and generalize the valid time range for the session necessary to complete the entire process of propagating messages from one node to another node.

This work makes the following contributions to e-business and social media analysis:

(1)A generalized suffix data structure (GST) is developed to support mining and querying sequential events in different applications, including: information propagation on social networks, frequent episodes mining, astronomical sequential pattern mining, and etc.

(2)Community-based algorithms can be integrated into the designed GST structure while discovering the propagation paths and influential nodes with various degrees of granularity of community.

(3)Integrating the above advantages, we provide an analysis tool that can help enterprises to discover the propagation paths and influential nodes at different levels of granularity of social networks in a given certain time period.

The rest of this paper is organized as follows. Section 2 reviews pertinent literature, while addressing both empirical and theoretical aspects of the role of social network analysis in information diffusion. Section 3 then introduces the proposed framework, which largely consists of system architecture and problem formulation of a social network. Next, Section 4 discusses the social features and metrics of the proposed method and presents the proposed classification approach, scheduling algorithm. Additionally, Section 5 show the experimental results. Conclusions are finally drawn in Section 6, along with recommendations for future research.

## 2. Related Work

Recently, to explore why and how fast does the message propagates over a social network, Berlingerio *et al.* [5] developed the so-called temporally annotated sequences mining algorithm combined with the concept of graph mining to find frequent sequential patterns. The proposed methodology can help in finding temporal behaviors together with the roles and characteristics of users in a social network. Tang *et al.* [6] designed a topical affinity propagation (TAP) approach and a topical factor graph model to deal with the problem of topic-level social influence propagation in social networks. To raise the scalability of TAP approach in real social networks, they also implemented a distributed-based TAP approach under the Map-Reduce framework. Wang *et al.* proposed a community-based greedy algorithm for mining top-K influential nodes. They first designed a community detection algorithm by taking into account information diffusion in a social network, and applied a dynamic programming algorithm for selecting communities to find influential nodes in mobile social networks [7]. However, these methods do not design a proper data structure for querying the influenced or influential nodes or paths in a given limited time.

Buccafurri and Lax applied the concept of probability to recover some form of transitive behavior in social networks, and integrated it into similarity-based methods to enhance the effectiveness of information dissemination [8]. Takeuchi *et al.* [9] utilized the similarity of propagation forms to extract the relationships of information on a social network. They showed that that the messages have duplicate links tend to be similar, and those have high rate of coincidence of the directions tend to be complementary. Although based on similarity-based methods for influence maximization problem in social networks, these works have seldom been explored to support analyzing social networks from different granularities of topic community.

Kimura *et al.* aimed to explore what kind of structure feature of social networks is correlated with the greedy algorithm, and demonstrated that the SR-community structure can be a better structure than the community structure introduced by Newman and Leicht [10, 11]. Anagnostopoulos *et al.* [12] applied statistical methods which are called the shuffle test and the edge-reversal test for identifying and measuring social influence on Flickr. Weng *et al.* [13] considered the homophily phenomenon in a community of Twitter, and designed a TwitterRank algorithm to measure the topic-sensitive influence of users in Twitter. Goyal *et al.* [14] utilized static probability models and time-dependent models to compute the success ratio of delivering information of a user.

From the viewpoint of social marketing, Domingos and Richardson viewed customers of market as nodes of a social network instead of a set of independent entities, and model the influence among their relationships as a markov random field to optimize marketing decisions for viral marketing [15, 16]. Cha *et al.* [17] attempted to observe how messages spread through social networks. They collect and analyze massive amounts of pictures that are disseminated in the Flickr to answer questions about information diffusion. The empirical results showed that contrast to the popular intuition of viral marketing; they found that even popular photos are spread locally and slowly.

However, all of the above-mentioned research has seldom been explored to provide a flexible data structure for discovering and querying influential nodes and paths in social networks. Therefore, this work employed the hierarchical-based community detection algorithm to classify users into different levels of topic community, and integrated it into the proposed generalized suffix algorithm to analyze influential nodes and paths from specification to generalization of topic communities in a given time period.

## 3. Overall Framework

This work attempts to design the analysis about social network of influence. Experiments on our methodology and system will be conducted based on queries of Digg's public data. The Digg is a social news website. Digg provide the API (Application Programing Interface) to the public [18]. In this section, we briefly introduce the basic structure of social network influence analysis. The analysis system is comprised of the three step: the data acquirement and processing, the generalized propagation suffix tree, and the request processing from the user interface. Step 1 use the Digg API to obtain the concept hierarchy from news categories of Digg, and take the news info of users to disseminate record to user documents, we then design a web crawler tool to obtain the Digg user follower relationship to social network construction. Step 2 then obtain user behavior analysis and find propagation paths to construction of the generalized propagation suffix tree, and step 3 provide query module of the query engine for users.

# 4. Methodology

## 4.1. User Behavior Analysis

We assume the context of an article represents the user's interest, and every article belongs to an interest community. Using the articles the user posted or reposted from other user in the past as references, we search for which community the user should belong to. After we classify all users to some communities in the lowest level of concept hierarchy, we can promote the point of view of the analysis from user to the more generalized community which the user belongs to. Thus we can analyze the message propagating patterns of the social network in different granularity.

As Figure 1 shows, if the user’s documents are mostly about IPhone, we will recognize that he is interested in IPhone, and classify him into the IPhone community in level 1 as well as Cell-Phone community in level 2.

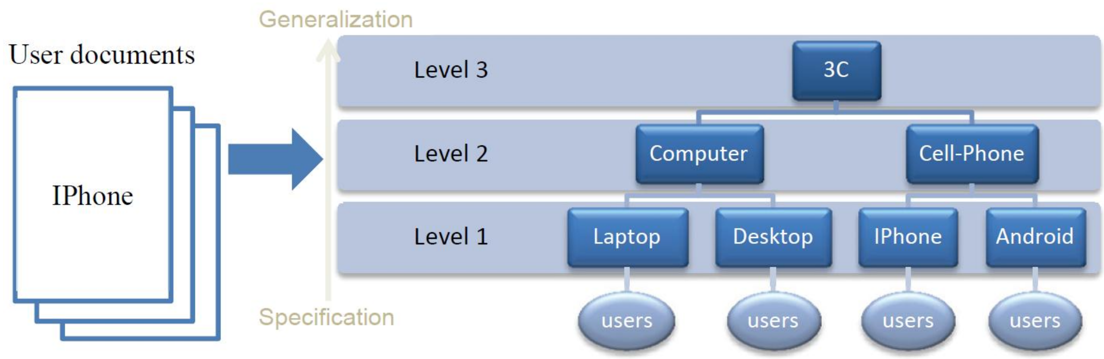



**Figure 1. User Behavior Analysis**

## 4.2. Find Propagation Path

In our research, we construct the social network by the subscribing relation in Digg. As Figure 2 shows, there is an Influence relation from User A to user B, which means A is subscribing B's articles; therefore A is a follower of B.

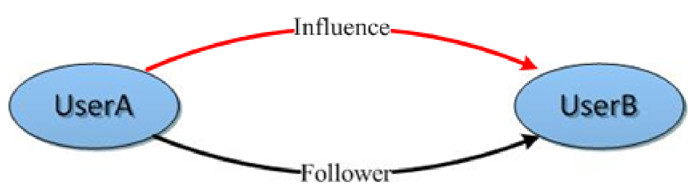



**Figure 2. Influence Relation**

The information of each document is shown as the format below. For each document, it has a record of its URL and the posted time. In Figure 3, User A posted a document having URL1 at the time 20, and User B posted the same document at time 22, then a propagation path from User A to User B is resulted. Figure 4 shows the following relationship graph of these users.

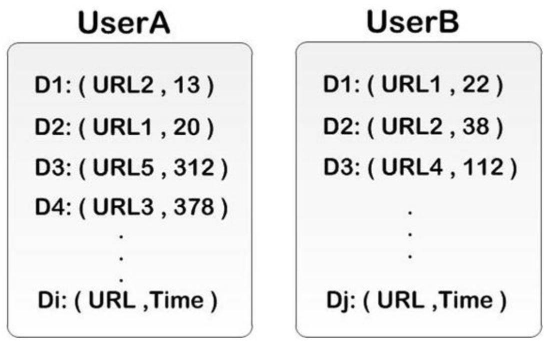



**Figure 3. Information Format**

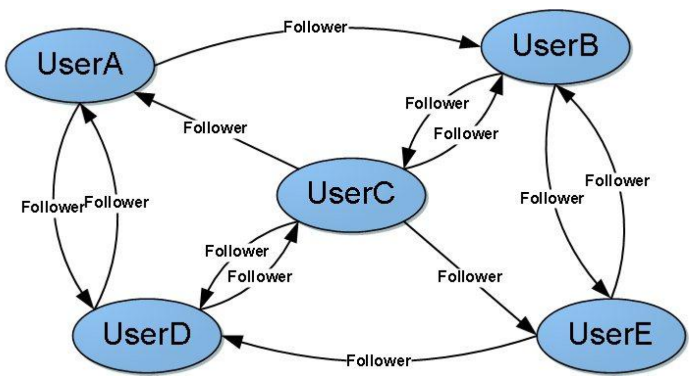
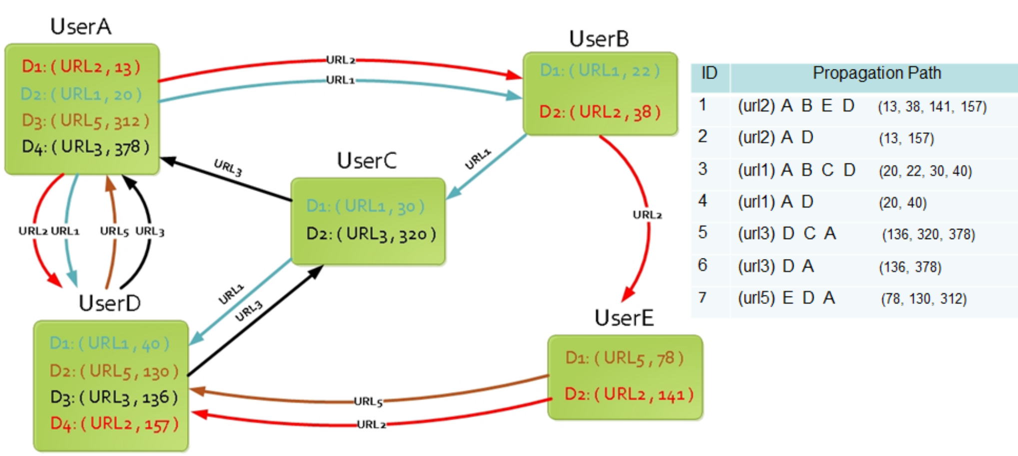



**Figure 4. Users Follower Relationship**

Figure 5 shows a propagating example. A document with URL2 was posted by User A at time 13; this document was later posted by the follower of User A (which is User B) at time 38, posted by the follower of User B (which is User E) at time 141, and posted by the follower of User E and User A (which is Used D) at time 157.

For the document with URL2, two propagation paths were created in the end, which are A->B->E->D and A->D.




| ID | Propagation Path | |
|---|---|---|
| 1 | (url2) A B E D | (13, 38, 141, 157) |
| 2 | (url2) A D | (13, 157) |
| 3 | (url1) A B C D | (20, 22, 30, 40) |
| 4 | (url1) A D | (20, 40) |
| 5 | (url3) D C A | (136, 320, 378) |
| 6 | (url3) D A | (136, 378) |
| 7 | (url5) E D A | (78, 130, 312) |

**Figure 5. Propagation Path**

In order to make sure the message propagation path can truly reflect the influence of propagation behavior, we set a threshold of propagation path session duration. If the same document is reposted by another user after a long time, we may consider cutting the propagation path as two separate paths, since the later user may not be influenced by the former; he may post this document because of other reasons. As the example shown in Figure 6, assuming the time threshold is 100 minutes, for the propagation path of document with URL2, the post time sequence is 13, 38, 141, 157, we cut the path between B->E since User E posted this document 103 minutes later than User B. This resulted two new propagation path. (A->B and E->D).

| ID | Propagation Path |
|---|---|
| 1 | (url2) A B E D (13, 38, 141, 157) |
| 2 | (url2) A D (13, 157) |
| 3 | (url1) A B C D (20, 22, 30, 40) |
| 4 | (url1) A D (20, 40) |
| 5 | (url3) D C A (136, 320, 378) |
| 6 | (url3) D A (136, 378) |
| 7 | (url5) E D A (78, 130, 312) |

| ID | Propagation Path |
|---|---|
| 1 | (url2) A B (13, 38) |
| 2 | (url2) E D (141, 157) |
| 3 | (url1) A B C D (20, 22, 30, 40) |
| 4 | (url1) A D (20, 40) |
| 5 | (url3) C A (320, 378) |
| 6 | (url5) E D (78, 130) |

**Figure 6. Propagation Path Session**

### 4.3. Generalized Propagation Suffix Tree

The generalized weighted suffix tree is derived from the concept of the weighted suffix tree [19]. The hierarchical structure of social networks is generated and integrated into the original weighted suffix tree to analyze different granularities of message propagation patterns.

According to Figure 6, the generalized suffix tree was constructed by the order of messages received time. And Figure 7 to 10 shows the generalized suffix tree was constructed by the propagation path. We create the tree nodes in the order of the path. For the path of URL 2(A->B), we first construct a root node, and insert the string AB, and insert another AB string but this time we delete its prefix. If we find the same node under the root while insertion, we will not create a new node.

Nodes on other path are created the same way as ID1. Each node will record the ID of the propagation path. More number of IDs recorded in the node means there's more appearance of this path. The number of different ID of propagation path is the weight value. Figure 10 shows the result of the tree after adding all message propagation paths. We define the propagation path session between each level 100 minutes. When transforming form user node to the communities, the tree node also records the first time for the message to propagate to a user in the corresponding community in level suffix tree (First propagation time), and the time for the message to propagate to all users in the corresponding community in level suffix tree (All propagation time).

| ID | Propagation Path | |
|---|---|---|
| 1 | (url2) A B | (13, 38) |
| 2 | (url2) E D | (141, 157) |
| 3 | (url1) A B C D | (20, 22, 30, 40) |
| 4 | (url1) A D | (20, 40) |
| 5 | (url3) C A | (320, 378) |
| 6 | (url5) E D | (78, 130) |

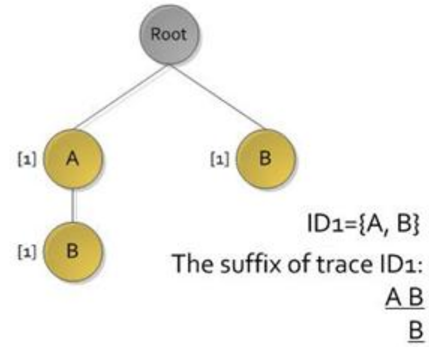



**Figure 7. The Suffix of Trace ID1**

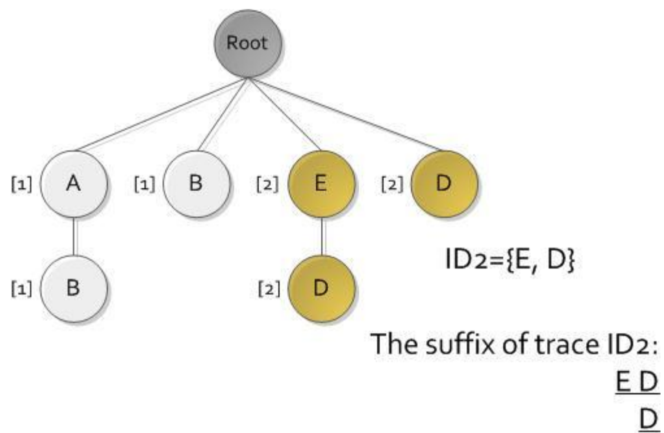



**Figure 8. The Suffix of Trace ID2**

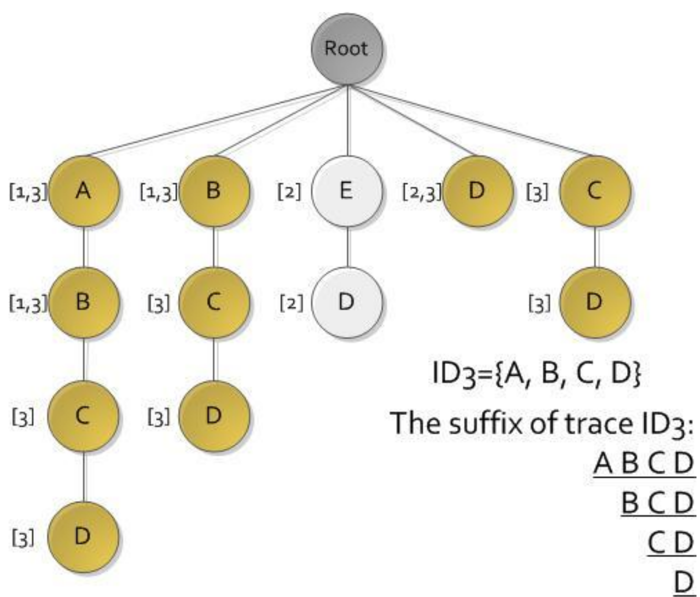



**Figure 9. The Suffix of Trace ID3**

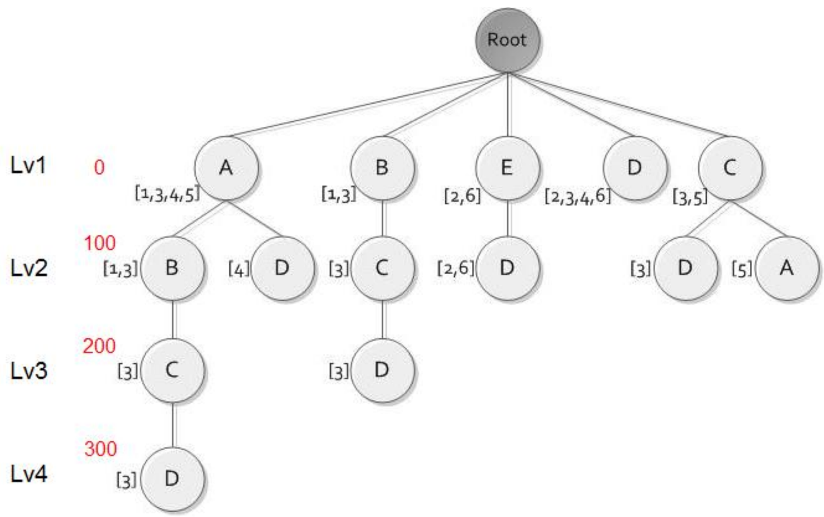
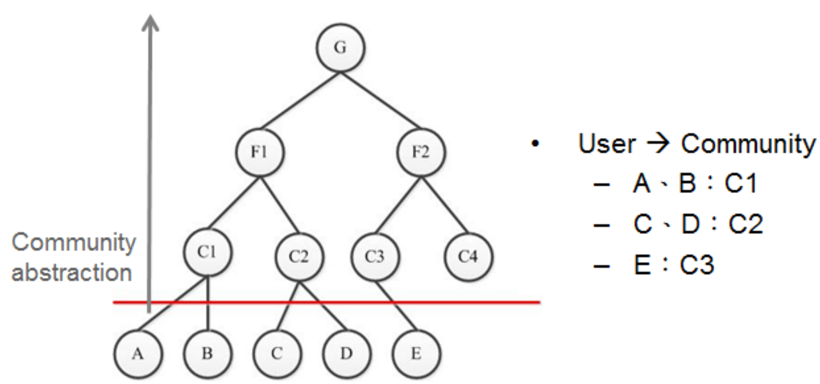



**Figure 10. Construction of Suffix Tree**

In Figure 11, as an example of Using concept hierarchy structure, example of let User A, B belongs to community C1, and User C, D belongs to C2; we improve the propagation path between users to propagation path between communities. For the path of ID3, we first change the user to the communities they belong to (A->B->C->D ➔C1->C1->C2->C2), then merge to the same community. So the community message propagation path of ID would become C1->C2.




| ID | Propagation Path |
|---|---|
| 1 | (url2) A B |
| 2 | (url2) E D |
| 3 | (url1) A B C D |
| 4 | (url1) A D |
| 5 | (url3) C A |
| 6 | (url5) E D |

| ID | Propagation Path |
|---|---|
| 1 | (url2) C1 C1 |
| 2 | (url2) C3 C2 |
| 3 | (url1) C1 C1 C2 C2 |
| 4 | (url1) C1 C2 |
| 5 | (url3) C2 C1 |
| 6 | (url5) C3 C2 |

| ID | Propagation Path |
|---|---|
| 1 | (url2) C1 |
| 2 | (url2) C3 C2 |
| 3 | (url1) C1 C2 |
| 4 | (url1) C1 C2 |
| 5 | (url3) C2 C1 |
| 6 | (url5) C3 C2 |

**Figure 11. Community Abstraction**

Figure 12 shows the transformation form user node to the communities, and it records the First propagation time and All propagation time.

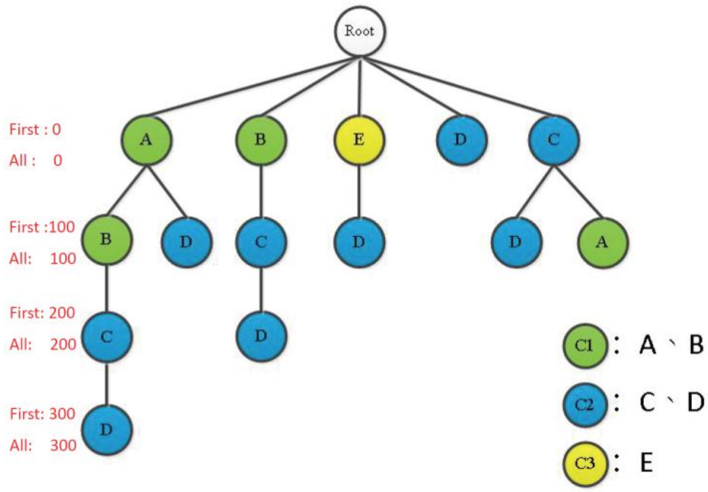



**Figure 12. The Node Generalization**

After node promotion, we merge the nodes in the same community in the propagation path of the trees; the merging algorithm works as Depth-First search:

1. Start form the root node.
2. Check the child nodes of the current node, if they belong to the same community as their parent, which means the propagation in the same community, then merge the child nodes. (Up-down merge)
3. Check the child nodes of the current node, if they belongs to the same community as their sibling, then merge the sibling node. (Left-right merge).
4. Each node perfume Up-down merge then Left-right merge, till every node have checked themselves.

In the process of merging, parent node is the one that merge the child node of another node; we save the "First reaching time" and the "All reaching time" of messages in each node. The ID list of all different propagation paths of the merged nodes are recorded and the path weight is the number of different paths.

Figure 13 is the suffix tree after node promotion. Firstly, the merge process begins at the root node, since root node is not a user node, we check if the child nodes are the same propagation path. There're 2 C1 child nodes and 2 C2 child nodes, the nodes at level 1 are equal to the starting point for the path, so we merge the same starting nodes to save unnecessary spaces. Secondly we check node C1 at level1, we merge the child node whom is also C1, which means the message propagates in the same community. After that the grandchild nodes of C1 node at level1 should connect to their grandparent node; for the merged node we pick the smallest First propagation time and largest All propagation time. Thirdly, we recheck if there's any child node of C1 is also C1, and merge them again. Since the child nodes are C2, these sibling node merge together and set the new time parameter. Figure 14 shows the final result generalized suffix tree after merging.

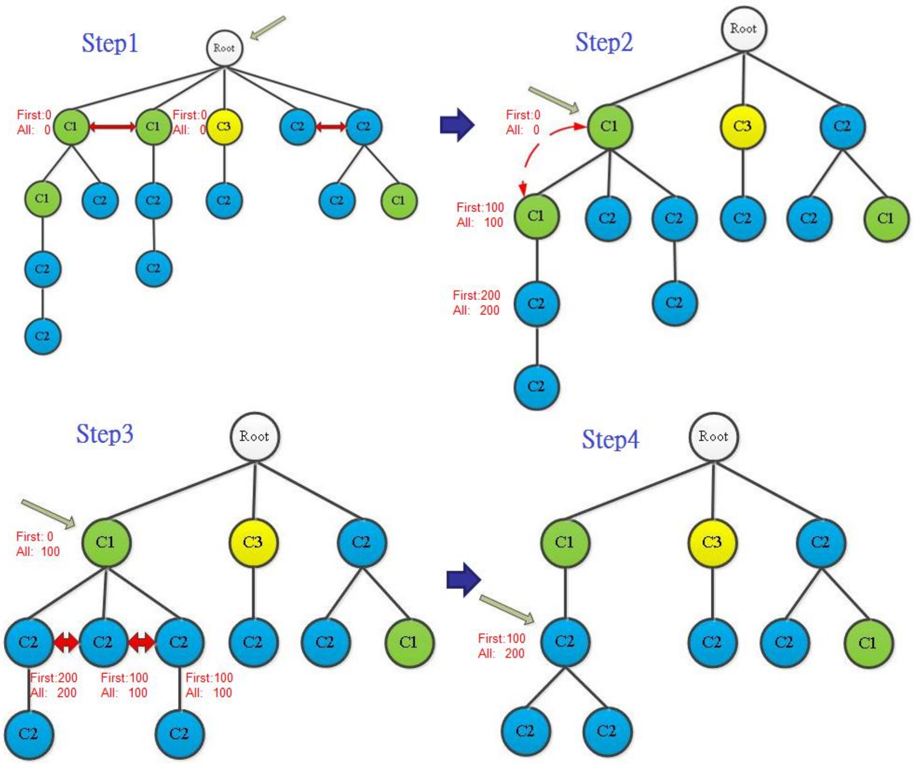



Figure 13. The Merge Step One

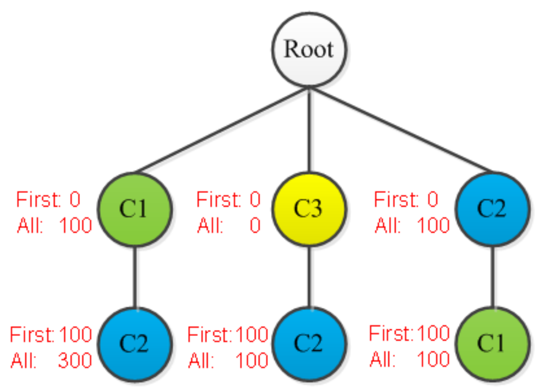



Figure 14. The Generalized Propagation Suffix Tree

### 4.4. Query Module

The analysis method of generalized propagating suffix tree we proposed in our research can help analyzing the message propagation pattern in different granularity of hierarchical communities. It provides users to mine and find out the most frequent propagation path, more influential communities or users, and the fastest time of propagating messages to a community and the time needed to spread the whole community, to detecting the power and time of influence of the communities.

We provide eight kinds of analysis queries for users to observe and analyze social networks as follows:

(1)Mining top-K frequent propagation paths under specific propagation length. We read all nodes in a specific length. By comparing the accumulated the frequency weights of nodes in corresponding length, the most frequent k paths can be selected.

(2)Mining top-K communities or users with most number of distinct paths. The answer is provided by calculating the number of distinct paths of the beginning parent node.

(3)Mining top-K users or communities which have the most distinct propagation path. We can find out who have propagated the message to most different users or communities by calculating how many different users or communities are there in the sub trees of the beginning parent node.

(4)Under specific time limitation, find top-K users or communities who can spread to most different users or communities. We find the users or communities who can propagate the messages to most different users or communities by calculating the sum of different users or communities in the sub trees of a beginning parent node under a time limitation.

(5)Find all the users who were influenced by a specific user in a certain time. The desired answer is calculated by summing up the number of different users or communities in specific time duration in the sub trees of a specific node.

(6)Find top-K users who are more likely to be influenced by a specific user. For this kind of queries, we seek for the child nodes which has greater path weight from a specific user or community.

(7)Find top-K users or communities who are likely to influence a specific user in a certain time. It is done by finding the parent nodes which have greater ID path weight.

(8)Find the most likely propagating and spreading time of the propagation from a community to another community. The desired information can be extracted from encoded time limit information in the corresponding nodes.

Note that all the previous queries can be computed with information directly from the suffix tree, without having to scan and access from the potential large dataset in the original database. With these eight queries, a salesman can find a frequent propagation path to send the advertisement effectively by focusing on those who are influential in the social network and suitable for endorsing the goods of the enterprise, or even hold a preferential activity in a period of time. With the strategies made by the support of the system, it can achieve the popularity and public praise of the enterprise efficiently.

## 5. Experiments

### 5.1. Dataset Collection and Preprocessing

The dataset are collected last for four months from Digg website [20]. Digg is a website for users to share information from anywhere on the web, such as online news, blog, photos, and videos. We collected 5,868 news, which were shared by 58,175 distinct users. The users of Digg website are classified into different levels of topic communities based on our proposed community classification method [4]. The number of Digg users is 12,717 after we filtered out users who shared news less than five. Table 1 shows the Digg website category classifies all news as 10 categories (which are level_2 on the hierarchical framework), and these 10 categories will also be classified as thinner 47 sub-categories (leve_1).

**Table 1. Digg Website Category [4]**

| | | |
|---|---|---|
| Business | Business | |
| Entertainment | Movies<br>Television<br>Music | Comic_and_Animation<br>Celebrity |
| Gaming | Nintendo(Wii)<br>Xbox<br>PlayStation | Gaming_Industry_News<br>Web_games<br>PC_games |
| Lifestyle | Travel_and_Places<br>Educational<br>Arts_and_Culture | Health<br>Autos<br>Food_and_Drink |
| Offbeat | Comedy<br>People | Pets_and_Animals<br>Other_Stuff |
| Politics | Political_Opinion | Politics_News |
| Science | Space<br>Environment | General_Science |
| Sports | Basketball<br>Extreme<br>Golf<br>Tennis<br>Other_Sports | Baseball<br>Hockey<br>Motorsport<br>Football |
| Technology | Microsoft<br>Security<br>Programming<br>Apple<br>Gadgets | Software<br>Hardware<br>Industry_News<br>Linux/Unix<br>Design |
| World News | World News | |

As Table 2 shows, we adopt three different session time threshold to mining the propagation behavior of users. 28,229 propagation paths were mined after one hour, and the longest propagation path length was 12. 120,513 paths were mined after 3 hours, and the longest path length was 15. 415,923 paths were mined after 6 hours and the longest path length became 18. Besides, we also save the record of mining 7,760,003 paths without any time limits.

**Table 2. Propagation Path**

| Session | Paths | Longest path |
|---|---|---|
| No Session | 7,760,003 | |
| 6 hours | 415,923 | 18 |
| 3 hours | 120,513 | 15 |
| 1 hour | 28,229 | 12 |

### 5.2. Results

All experiments in this work are performed on a machine with Intel Core 2 Duo 2.0 GHz CPU, 4 GB RAM, and run on Microsoft Windows 7 Professional. Table 3 shows number of nodes, maximum level of tree, and construction time of generalized weighted suffix tree separately under three different session time.

**Table 3. The Generalized Propagation Suffix Tree of Users**

| Session | Nodes | Max Level | Time of construct |
|---|---|---|---|
| 6 hours | 1,067,483 | 18 | 2 hours and 10min |
| 3 hours | 256,887 | 15 | 28min |
| 1 hour | 29,683 | 12 | 3min |

Table 4 shows the merging time at different levels of the suffix tree data structure under three different session time.

**Table 4. The Generalized Propagation Suffix Tree of Community**

| Level | Session | Nodes | Max Level | Time of merge |
|---|---|---|---|---|
| Level_1 | 6 hours | 3602 | 13 | 36140s |
| Level_1 | 3 hours | 2086 | 13 | 1247s |
| Level_1 | 1 hour | 1081 | 11 | 11.852s |
| Level_2 | 6 hours | 1805 | 13 | 1430s |
| Level_2 | 3 hours | 1005 | 13 | 72s |
| Level_2 | 1 hour | 458 | 11 | 0.752s |

This research proposed a novel hierarchical message propagating path mining method to observe the behavior of users in a social network. In the past work it took a great time to sieve the the data from a huge database to do the analysis of the influence of users in the community. In our work we combine the tree structure with data mining technique, and combine the influence and message propagation method with concept hierarchy, to construct the generalized propagation suffix tree. With this suffix tree, we can acquire the data in a short period of time, and even promote the suffix tree to a community level from the user level without constructing another new suffix tree. This analysis variety includes time limitations, first time to propagate, spreading time. It provides the user of this system to find out the most influential users or communities, the most frequent propagation path, and it helps them making efficient marketing plans. With the information obtained from this system, the marketing department could find the most valuable marketing strategy to help the enterprise costing down to achieve the most benefit.

## 6. Conclusions

This research proposes a novel generalized weighted suffix data structure for querying message propagation behavior of users at different levels of granularity of social networks. With this suffix tree, we can observe and analyze how information diffusion in massive online social networks in a short period of time, and apply it for different applications, such as viral marketing and target marketing. In addition, the designed generalized weighted suffix structure will be applied to different fields in the future, such as gamma-ray burst mining and analysis in astronomy, intrusion detection in computer security, *etc.*.

## Acknowledgements

The authors would like to thank the National Science Council, Taiwan, for financially supporting this research under the Contract No. NSC 99-2219-E-002-021.

## References

[1] eMarketer Digital Intelligence : Facebook Drives US Social Network Ad Spending Past $3 Billion in 2011, http://www.emarketer.com/Article.aspx?R=1008180, **(2011)**.

[2] Magnani M, Montesi D and Rossi L, “Information propagation analysis in a social network site”, Proceedings of the 2010 International conference on Advances in Social Network Analysis and Mining, pp. 296-300, **(2010)** Aug 9-11; Odense, Denmark.

[3] Chen VHH and Duh HBL, “Investigating User Experience of Online Communities: The Influence of Community Type”, Proceedings of the 2009 International Conference on Computational Science and Engineering, pp. 509-514, **(2009)** Aug 29-31; Vancouver, Canada.

[4] Wang MF, Kuo YL, Tsai MF, Tang CH and Huang K, "Hierarchical Role Classification based on Social Behavior Analysis", Proceedings of the 8th International Conference on Advances in Mobile Computing and Multimedia, pp. 426-429, **(2010)** November 8-10; Paris, France.
[5] Berlingerio M, Coscia M, Giannotti F, Monreale A and Pedreschi D, "Foundations of Multidimensional Network Analysis", Proceedings of the 2011 International Conference on Advances in Social Networks Analysis and Mining, pp. 485-489, **(2011)** July 25–27; Kaohsiung, Taiwan.
[6] Tang J, Sun J, Wang C. and Yang Z, "Social Influence Analysis in Large-Scale Networks", Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 807-816, **(2009)** June 28 - July 1; Paris, France.
[7] Wang Y, Cong G, Song G and Xie K, "Community-based Greedy Algorithm for Mining Top-K Influential Nodes in Mobile Social Networks", Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1039-1048, **(2010)** July 25-28; Washington, USA.
[8] Buccafurri F and Lax G, "Improving Similarity-Based Methods for Information Propagation on Social Networks", Proceedings of the 2010 Networked Digital Technologies Communications in Computer and Information Science, pp. 391-401, **(2010)** July 7-9; Prague, Czech Republic.
[9] Takeuchi S, Kondoh T, Akiyoshi M and Komoda N, "Evaluation of Similarities of Propagation Forms on Social Network for Extracting Relationships of Information", Proceedings of the SAINT 2008, IEEE Computer Society, pp. 341-344, **(2008)** July 28 - Aug 1; Turku, Finland.
[10] Kimura M, Saito K, Nakano R and Motoda H, "Extracting influential nodes for information diffusion on a social network", Proceedings of the 22th AAAI Conference on Artificial Intelligence, pp. 1371-1376, **(2007)** July 22–26; Vancouver, Canada.
[11] Kimura M, Yamakawa K, Saito K and Motoda H, "Community Analysis of Influential Nodes for Information Diffusion on a Social Network", Proceedings of the 2008 International Joint Conference on Neural Networks, pp. 1358-1363, **(2008)** June 1-6; Hong Kong, China.
[12] Anagnostopoulos A, Kumar R and Mahdian M, "Influence and Correlation in Social Networks", Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 7-15, **(2008)** Aug 24-27; Las Vegas, USA.
[13] Weng J, Lim EP, Jiang J and He Q, "TwitterRank: finding topic-sensitive influential twitterers", Proceedings of the 3th ACM international conference on Web search and data mining, pp. 261-270, **(2010)** Feb 4-6; New York, USA.
[14] Goyal A, Bonchi F and Lakshmanan LVS, "Learning Influence Probabilities in Social Networks", Proceedings of the 3th ACM international conference on Web search and data mining, pp. 241-250, **(2010)** Feb 4-6; New York, USA.
[15] Domingos P and Richardson M, "Mining the network value of customers", Proceedings of the 7th International Conference on Knowledge Discovery and Data Mining, pp. 57-66, **(2003)** Aug 24-27; Washington, USA.
[16] Richardson M and Domingos P, "Mining knowledge-sharing sites for viral marketing", Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 61-70, **(2002)** Jan.28 - Feb.1; Nara, Japan.
[17] Cha M, Mislove A and Gummadi KP, "A Measurement-driven Analysis of Information Propagation in the Flickr Social Network", Proceedings of the 18th international conference on World Wide Web, pp. 721-730, **(2009)** April 20-24; Madrid, Spain.
[18] Digg developers (Application Programing Interface), http://developers.digg.com/documentation.
[19] Wang MF, Wu YC and Tsai MF, "Exploiting frequent episodes in weighted suffix tree to improve intrusion detection system", Proceedings of the 22th International Conference on Advanced Information Networking and Applications, pp. 1246-1252, **(2008)** March 25-28; Okinawa, Japan.
[20] Digg website, http://www.digg.com.
[21] Paik J, Nam J and Kim UM, "Finding Frequently Occurred Tree Patterns without Candidate Subtrees Maintenance", International Journal of Database Theory and Application, vol. 2, no. 1, **(2009)**, pp. 11-22.
[22] Shahriar MS and Anam S, "Towards Data Quality and Data Mining Using Constraints in XML", International Journal of Database Theory and Application, vol. 2, no. 1, **(2009)**, pp. 23-30.
[23] Mustafa A, Akbar A and Sultan A, "Knowledge Discovery using Text Mining: A Programmable Implementation on Information Extraction and Categorization", International Journal of Multimedia and Ubiquitous Engineering, vol. 4, no. 2, **(2009)**, pp. 183-188.