Message Spreading Model over Online Social Network with Multiple Channels and Multiple Groups

Sungmin Hwang and Kyungbaek Kim
Dept. Electronics and Computer Engineering
Chonnam National University
Gwangju, Republic of Korea
e-mails: {sungmin1511@gmail.com, kyungbaekkim@jnu.ac.kr}

***Abstract*—Understanding the characteristics of message spreading over online social network is important for estimating the influence of message initiated from arbitrary users. Past researches present some models, such as independent cascade model and linear threshold model, to explain the message spreading. Recent studies show many variations of previous models focused on different issues. In this paper, we focus on multiple channels that are used for communicating with each other, and multiple groups that react differently to the message coming from each channel, in order to observe a more detailed aspect of message spreading, such as spreading speed or chances to accept the message. Considering these properties, we propose a new message spreading model that has multiple member groups and multiple channels. We examine the impact of channel and group preference in message spreading by conducting extensive simulations of our suggested model. Through the simulations, we observed that considering multiple channels and multiple groups explains the speed and the coverage of message spreading in more detail.**

Keywords-Message Spreading; Online Social networks; Multiple Channels; Multiple Groups.

I. Introduction

Observing information diffusion has always been an interesting research area, and, thanks to the Internet, there are more researches coming out, which are related to maximizing the effect of viral marketing [1][2][5], or social influence in Social Network Services [3]. To understand the message spreading in social networks, many models have been suggested and some of them are frequently discussed [1]. Based on each model, many related researches have been done which focus on different aspects of social network such as finding the source of information [6], or finding effective way to spread information [1][2][4]. Also, there are newly suggested models for different attributes and goals [7][8]. We focus on the means which people use in the social network to communicate with others.

With the help of technologies developed recently, the number of ways which people use to communicate with each other is increasing. We call these ways to communicate as channels. Since members of social networks have many choices to send messages, we need to consider the properties of channels to study about message spreading speed. For the cases that require urgent message delivery, like accident or disaster aware services, not only coverage, but also speed of message spreading is an important property. Considering speed, every channel has its own unique properties, such as time it takes to send message or time it takes until receiver checks the message. Texting and phone calls can be examples. Since texting requires typing and cannot be sure whether the receiver checks the message instantly, it has longer time expectation compared to phone call which makes the receiver react instantly. For the message that has time constraints, or to observe the speed of message spreading, considering these properties is needed to properly estimate the diffusion.

While considering channels, we find one more thing to think about, namely, the preference of channel. Channel preference can be different from one user to another. According to this, considering each individual user for applying channel preference is encouraged, but it is practically impossible. As an alternative, we considered a user group that shares similar properties, such as age, income and professions, and the user group explains the characteristics and the preference of channels.

After grouping members in network, their common behavior can be considered. Focused on message spreading, we considered what channels they prefer when sending messages, and from what channel they accept the message and resend to others. As an example, teenagers will prefer using Short Message Service (SMS) [12] or instant messenger to send messages and may have higher chance to accept messages through these channels compared to messages from other channels. When communicating with same groups, this is not an issue. But, when communicating with members in other groups, this can bring a different aspect of diffusion due to their different preference of channels.

In this paper, we propose a new message spreading model which considers the properties discussed above such as multiple channels and multiple groups. In this model, a member who receives the message reacts differently by which channel the sender used and which group the member belongs.

The rest of this paper is structured as follows. In Section 2, related works of message spreading are explored, such as linear threshold model and independent cascade model. Section 3 describes the structure of our proposed model and how this model works related to real situations. In Section 4, we evaluate our model with the real-world social network data, and we conclude in Section 5.

II. Related Works

There are many models that explain how the information diffusion happens in social networks and two famous models are independent cascade model and linear threshold model. Each model shows a different aspect of influence.

A. Linear Threshold Model

Linear Threshold (LT) Model [1] describes the activation of a node as following the neighbors' major opinion or behavior. In this model, a social network is modeled as directed graph $G = (V, E)$, where the vertices of V represents individuals in the network and edges in E represents relationships and direction of an edge shows who is influenced by whom. Every node *v* in G is in state of either active or inactive and can only be activated once. Also, node *v* chooses a random threshold θ_v that has range of [0, 1].

A node *v* is influenced by each neighbor w according to the edge weight $b_{v,w}$ such that

$$\sum_{w \text{ neighbor of } v} b_{v,w} \leq 1 \qquad (1)$$

If the total weight of activated neighbors reaches the threshold θv, the node gets activated and affects other inactive nodes. Being affected by neighbor nodes explains the tendency of adoption of message or product when other neighbors already adopted it. A man will feel like buying something if many of his coworkers already have one. Some researches modified this model to explain product adoption in social network [5].

B. Independent Cascade Model

LT model expresses the diffusion process well, but does not fit into our purpose since its main idea lies in tendency of majority. So, we used the other model that is based on probability, the Independent Cascade (IC) Model [1]. Fig. 1 illustrates the activation process of the IC model. From the initial stage of the diffusion process, every node starts with an inactive state except the nodes in initial node set *A0*. A node can only be activated once, and when it is activated, it has one chance to activate neighbor nodes. In Fig. 1, when node *v* becomes active by the node *x*, it tries to activate each

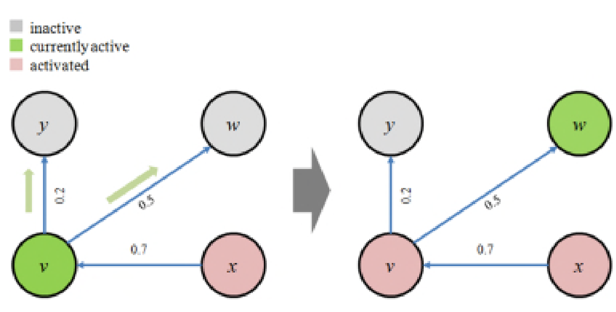

Figure 1. Activation process of Independent Cascade Model

neighbor node *w* and *y* which are in inactive state. By the chance $P_{v,w}$, which is 0.5, *v* becomes active. If node *v* fails to activate node *w*, node *v* cannot try to activate node *w* again. Once node *w* is activated, it can no longer be activated again and stays in activated state. The process continues until there is no further chance to activate neighbor nodes from each node. Unlike the LT model, this model describes the diffusion process as cascading of independent decisions made by each node. Since a node making own decision better fits our idea of separating channels and groups, we modify this model to explain the message spreading over social network, which considers multiple channels and multiple groups.

C. Considering Groups in Social Network

Classifying groups in social networks has been researched in order to improve the performance of services in social networks, such as language learning [9][10].

III. Message Spreading Model with Multi-Cahnnel Multi-Group

A. Modeling

To consider channel attributes and preferences, we modify the IC model discussed in Section 2. In our new model, a node *v* belongs to one of groups in the network based on grouping rules, which can be profession, age, or whatever that member of each group can share same channel preference. Each user group g_i, has distinguishing channel preference S_{ij} for each channel c_j to send messages,

$$S_{m\times n} = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1n} \\ S_{21} & & & \vdots \\ \vdots & & \ddots & \\ S_{m1} & & \cdots & S_{mn} \end{bmatrix} \qquad (2)$$

where m is the number of groups and n is the number of channels. Since a node *v* must select at least one channel, $\sum_{j=1}^{n} S_{ij} = 1$. It also has acceptance A_{ij}, the chance to accept and resend the message that comes through channel c_j,

$$A_{m\times n} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & & & \vdots \\ \vdots & & \ddots & \\ A_{m1} & & \cdots & A_{mn} \end{bmatrix} \qquad (3)$$

Here, $0 \leq A_{ij} \leq 1$, since acceptance of each channel works independently. A message sent through c_j has time delay d_j, which is determined by the channel used for activation.

Fig. 2 illustrates what happens in the model when a node *v* tries to activate its neighbor node *w*. When the process starts, node *v* decides which channel to use from c_1 to c_n based on S_{ij}, the channel preference of the user group it belongs. After the node *v* chooses channel c_j, *v* tries to activate a neighbor node *w* through that channel. After the time delay d_j, the time it takes a message to reach its target,

 ISBN: 978-1-61208-361-2

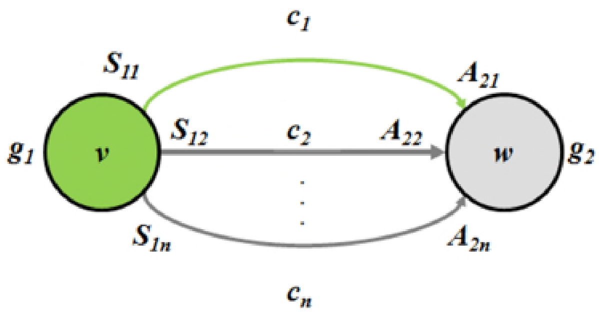

Figure 2. Channel selection and acceptance

the neighbor node w determines activation by the acceptance of the channel c_j, A_{ij}. With this individual activation process, Algorithm 1 in Fig. 3 shows the entire node activation process.

B. Correspondence between groups

Intra-group communication, which takes place between members of the same group may differ from that of inter-group communication. People's tendency of relying on the most preferred channel is one of the reasons that change how people react to information from certain channels. Considering correspondence between members of the same group, each member will have higher chance to accept the information from the channel they prefer most. This leads to the assumption that people may show higher acceptance to the information coming from the channel they prefer.

Algorithm 1 Node activation algorithm

set time=0
for each node $v \in AO$ **do**
  add v to activation queue Q
  (Q is timestamp priority queue)
  set T_v = time
end for
while (Q is not empty)
  if time = T_f, timestamp of first node in Q do
    set state of v activated
    for each neighbor w
      generate random number X,
      $0 < X \leq 100$
      set i = group of v
      for each channel c_j
        if $X > \sum_k^{j-1} S_{ik}$ and $X \leq \sum_k^{j} S_{ik}$ **then**
          set edge $E_{vw} = c_j$
      end for
      generate random number Y,
      $0 < Y \leq 100$
      set i = group of w
      if $Y \leq A_{ij}$ **then**
        set timestamp of w = time + d_j
        add w to Q
    end for
  else time++
end while

Figure 3. Node activation algorithm

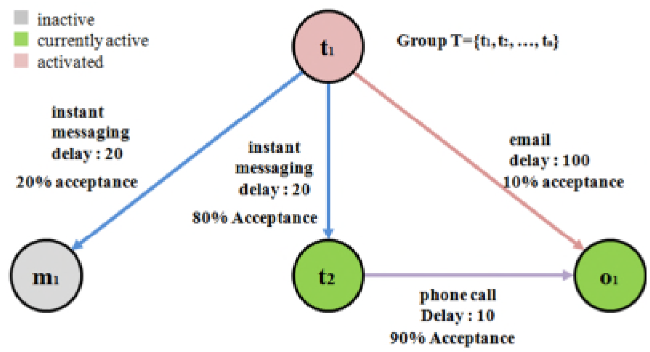

Figure 4. An example of MCMG model

TABLE I. CHANNEL PREFERENCE OF GROUPS

Channels	Channel selection			Acceptance		
	Instant messaging	*Phone call*	*Email*	*Instant messaging*	*Phone call*	*Email*
Group T	0.5	0.2	0.3	0.8	0.15	0.2
Group M	0.3	0.5	0.2	0.3	0.6	0.3
Group O	0.1	0.8	0.1	0.1	0.9	0.1

Fig. 4 shows an example of the activation process of the Multi-Channel Multi-Group (MCMG) model that has three groups and three channels. Based on the assumption that intra-group acceptance is strong and people rely on what they use most, each group shows highest acceptance on the channel which has highest selection chance. Following the group channel selection rate in Table 1, member t_1 in group T shows highest usage of instant messaging channel. After member t_1's attempt to activate inactive member m_1, t_2, and o_1, each member accepts or ignores the message based on acceptance chance in Table 1. t_2 shows 80% acceptance for the message through instance messaging channel while m_1 shows lower acceptance of 20% for the message from same channel. Group O as a group of members over age of 70, shows lower acceptance on email since it is possible that they cannot check the email at all but they show high acceptance in classical way of communication such as phone call. The reaction difference to each channel shows why we need to consider groups more seriously when we consider channels.

C. Simplified model of MCMG model

In some of previous researches, simple Independent Cascade Model that has single channel and single user group has been used to get the result of experiment due to calculation time and complex algorithm when considering complex model. A simplified version of our MCMG model is needed for those reasons, and for comparing MCMG model's aspect of message spreading with simple Independent Cascade Model. To simplify the model, we consider simplifying channels into one single channel that presents all other channels.

$$C = \{c_1, c_2, \cdots, c_n\} \qquad C \to c \tag{4}$$

$$\text{delay } D = \{d_1, d_2, \cdots, d_n\} \qquad D \to d_{avr} \tag{5}$$

The average delay of channel becomes delay of the simple channel. Also, due to the simplifying channels, preference of channels changes into

$$S_{m\times n} \to S_{m\times 1} \qquad A_{m\times n} \to A_{m\times 1} \tag{6}$$

where preference $S_{m\times 1}$ and $A_{m\times 1}$ has value of average preference for every channel.

Simplifying groups into one whole group should consider preferences conversion that fits to one group.

$$G = \{g_1, g_2, \cdots, g_m\} \qquad G \to g \tag{7}$$

$$S_{m\times n} \to S_{1\times n} \qquad A_{m\times n} \to A_{1\times n} \tag{8}$$

where simplified preferences

$$S_n = \text{avr}\{S_{1n}, S_{2n}, \cdots, S_{mn}\}\text{ , and} \tag{9}$$

$$A_n = \text{avr}\{A_{1n}, A_{2n}, \cdots, A_{mn}\} \tag{10}$$

show general channel selection and acceptance when group ratio in the networks are some.

If group ratio is considered,

$$R_{1\times m} = [R_1 \quad R_2 \quad \cdots \quad R_m] \tag{11}$$

$$RS_{mcmg} = S_{simplified} \tag{12}$$

where R is the group ratio.

This model simplifies all nodes into one group and all channels into single channel that has general tendency. The model is used for comparison in evaluation.

IV. Evaluation

A. Setup

1) Dataset

We use real world graph data set from Slashdot [11] to simulate our MCMG model. Slashdot is a technology-related news website known for user-submitted technology oriented news and it has Slashdot Zoo feature, which allows users to tag each other as friends or foes. We use this friend/foe links between users of Slashdot as our dataset which is obtained in November 2008.

Every node in the graph has its unique ids and randomly assigned to one of groups based on ratio of the groups. The first message starts from the fixed initial node.

TABLE II. Network Graph Statistics

Dataset statistics	
Nodes	77360
Edges	905468
Diameter(longest shortest path)	10
90-percent effective diameter	4.7

TABLE III. Channel Selection Probabilities of Groups

Channel Selection	Channel1	Channel2	Channel3	Channel4
g_1	0.70	0.20	0.5	0.5
g2	0.15	0.60	0.15	0.10
g3	0.10	0.40	0.40	0.10
g4	0.10	0.15	0.60	0.15
g5	0.10	0.10	0.10	0.70

TABLE IV. Acceptance Probabilities of Groups

Acceptance	Channel1	Channel2	Channel3	Channel4
g1	0.4	0.10	0.5	0.5
g2	0.10	0.40	0.10	0.5
g3	0.5	0.25	0.25	0.5
g4	0.5	0.5	0.40	0.5
g5	0.5	0.5	0.10	0.40

We consider 5 groups,

$$G = \{g_1, g_2, g_3, g_4, g_5\}$$

and delays of channel, D, where

$$D = \{80, 150, 300, 450\}$$

to see the variations of channels and groups. Delay differences between channels show the time differences caused by sending and receiving. As an example, email usually has long delay due to their time taken from sending and checking the message while calling on a phone has relatively short delay. We consider 4 channels in this case.

Table 3 and Table 4 show the probabilities of channel selection and acceptance, respectively. Each group has been set to have higher chances of acceptance to the channels they show more selection chances, due to the assumption we have made in modeling.

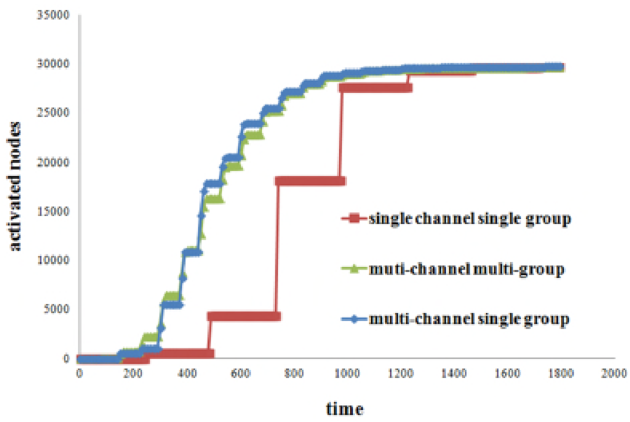

Figure 5. Activation progress of MCMG model and simplified model

2) Simulatior

We implemented a simulator that examines activation of nodes per time based on the proposed model. It starts from a set of initial nodes and tries to activate the neighbor nodes. When the node is successfully activated, the status of node changes to activated and the node is stored in priority queue with time stamp value of current time plus delay of channel used. When the time reaches the value of time stamp, the node tries activation of its neighbor node and deleted from the queue after the process. Time starts from 0 and keeps increasing until there is nothing in the queue. The result contains time and total activated nodes at that time.

B. Result

1) Multiple channels

Fig. 5 shows the coverage of message spreading with 3 different models, as a function of time. "multi-channel multi-group" represents the proposed model. "multi-channel single group" represents a simplified model which merges all groups into a single user group. "single channel single group" means the simplified model described in section 3.C. In this evaluation, the percentage of each group is set to equal, 20%. As shown in Fig. 5, cases using multiple channels show similar result, the fluent curves. Multiple choices a node can make brought diversity of time delays that causes different progress in diffusion. But the case that has only single channel shows step shaped line due to lag of variance in delay. Though it has same coverage (the total number of activated nodes), it shows different speed in the middle of spreading message.

This can be critical under the circumstances that have time constraint, or that require intermediate values since those cases require spreading speed at the certain point of time. Short period marketing, such as limited sale policy, can use the proposed model to get proper analysis of effect. Also, analyzing warning message spreading such disaster-aware service, which is time sensitive, requires activation rate at certain point too.

The simplified model (single channel single group) is not appropriate for observing detailed progress of diffusion. It might save the calculation time to conduct message spreading process and to find the eventual coverage, but it does not consider the diversity that it can easily miss few important characteristics of channels. Also as the IC model, the MCMG model surely follows the concept 'independent' since every node decides what channels to use on its own.

2) Multiple groups

To observe the impact of groups to message spreading, we tried different user ratio for each groups. Fig. 6 shows the result of considering group ratio in network. In Fig. 6(a), group1 is the major group which prefers fastest channel such as internet phone and SNS instant messaging. In Fig.

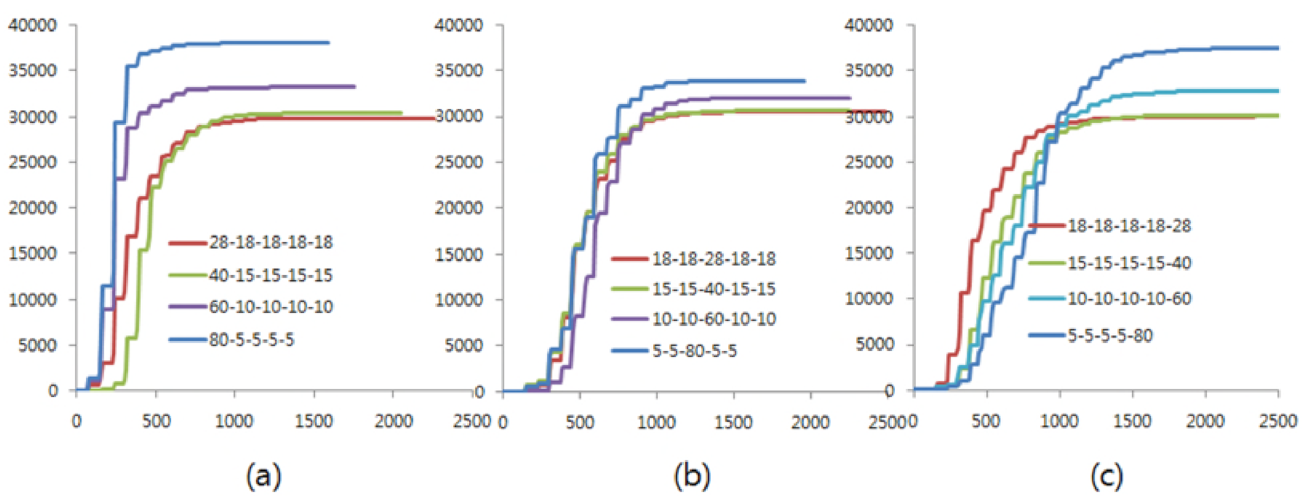

Figure 6. Message spreading under various group ratios. Number in each legend represent percentage of groups by sequence

6(c), major group is group5 which prefers slowest channel such as SNS posting (not instant messaging). In Fig. 6(b), major group is group 3 whose preference is in the middle between group 1 and group 5. In Fig. 6(a), as the percentage of group 1 increases, the speed of message spreading increases. On the other hand, in Fig. 6(c), as the percentage of group 5 increases, the speed of message spreading decreases. That is, the speed of message spreading depends on which channel the major group prefers.

Besides the speed of message spreading, there is one more issue to discuss, the coverage. In Fig. 6(a) and Fig. 6(c), it is observed that the coverage of message spreading increases as the percentage of major group increases. It is because that the intra-group acceptance set higher than other in the current evaluation setting. With the same reason, when the major group is group 3 which prefers two channels, the coverage is less than other two cases like Fig. 6(b). These evaluation results show that the coverage of message spreading may be highly affected by the ratio of user group and preferences of groups.

These results lead us believe that considering groups is an important issue when those groups have their distinguishable channel preference. Since online social network in these days consist of many groups with unique attributes of their own, examining channels they prefer can help us to follow complex message spreading in the network. Also, predicting message spreading without considering groups can cause an overestimation or underestimation compared to which considers groups.

V. CONCLUSION AND FUTURE WORK

Nowadays, the number of channels people use for communication is increasing with the fast development of messenger programs and technology. Google Talk, Skype, or Face Time can be the examples. These channels have unique characteristics that should be considered when observing message spreading over social networks. In addition, every social networks and groups in the network have preferred channel of their own and this can make the relation between channels and groups. The key point is that each group reacts differently to message sent through each channels. This affects not only the speed of message spreading, but also coverage of the message. For that reason, both channel and group in social network should be considered to examine message spreading.

In this paper, we proposed a new message spreading model over online social network which considering multiple channels and multiple groups. With multiple channels, we considered delay and preference of channels, and we found meaningful results from dividing channels. We expect considering the other properties of channels also will lead to more detailed/accurate aspect of message spreading. Likewise, more detailed grouping policy will be helpful to get various results.

As a natural extension of this work, our future work is finding relations between channels and groups and applying them into information diffusion model, which are important to find more effective ways to spread certain messages over online social network.

ACKNOWLEDGEMENTS

This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (NIPA-2014-H0301-14-1014) supervised by the NIPA(National IT Industry Promotion Agency).

REFERENCES

[1] D. Kempe, J. Kleinberg, and E. Tardos, “Maximizing the spread of influence through a social network,” Proc. ACM SIGKDD 2003, pp. 137-146.

[2] W. Chen, Y. Wang, and S. Yang, “Efficient influence maximization in social networks,” Proc. ACM SIGKDD June. 2009, pp. 199-208.

[3] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts, “Everyone’s an influencer: quantifying influence on twitter,” Proc. ACM international conference on Web search and data mining, 2011, pp. 65-74.

[4] A. Najar, L. Denoyer, and Patrick Gallinari, “Predicting information diffusion on social networks with partial knowledge,” Proc. 21st international conference companion on World Wide Web, April. 2012, pp. 1197-1204.

[5] S. Bhagat, A. Goyal, and L. V.S. Lakshmanan, “Maximizing product adoption in social networks,” in Proc. the fifth ACM international conference on Web search and data mining, February. 2012, pp. 603-612.

[6] T. Lappas, E. Terzi, D. Gunopulos, and H. Mannila, “Finding effectors in social networks,” Proc. the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, July. 2010, pp. 1059-1068.

[7] J. Yang, and J. Leskovec, “Modeling information diffusion in implicit networks,” Proc. the IEEE International Conference on Data Mining, 2010, pp. 599-608.

[8] X. Song, Y. Chi, K. Hino, and B. L. Tseng, “Information flow modeling based on diffusion rate for prediction and ranking,” Proc. the 16th international conference on World Wide Web, May. 2007, pp. 191-200.

[9] C. Troussas, M. Virvou, J. Caro, and K. J. Espinosa, “Language learning assisted by group profiling in social networks,” iJET, 8(3). 2013, pp. 35-38.

[10] C. Troussas, M. Virvou, J. Caro, and K. J. Espinosa, “Multivariate clustering for group language learning in Facebook,” IIMSS, 2013, pp. 80-88.

[11] Slashdot. [Online]. Available from http://slashdot.org [accessed May 2014]

[12] A. Stoica, Z. Smoreda, and C. Prieur, “A Local Structure-Based Method for Nodes Clustering: Application to a Large Mobile Phone Social Network”, The Influence of Technology on Social Network Analysis and Mining, Lecture Notes in Social Networks 6, pp. 157-184, 2013.

 ISBN: 978-1-61208-361-2