International Journal of Advanced Information Science and Technology (IJAIST) ISSN: 2319:2682
Vol.3, No.11, November 2014 DOI:10.15693/ijaist/2014.v3i11.19-26

Information Diffusion Models and Algorithms for Influence Maximization in Social Networks- A Survey

Lal Vikram Singh
Department of Computer Science
Pondicherry University, India

S.Siva Sathya
Associate Professor/Department of CSE
Pondicherry University, India

***Abstract*— Social networks are important mediums for spreading information, thoughts and influences among individuals. Social influence is the behavioural change of a person because of the perceived relationship with other people, organizations and society in general. In particular, the rapid growth of online social networks such as Facebook, Twitter and Google+ has intensified interests in this field. Several models have emerged in the recent days to conceptually model the spread of information in a social network. Also the most influential nodes to maximize the information spread has been studied extensively. In this Survey, the goal is to provide the readers with a comprehensive review of widely used theoretical diffusion models in social networks. The subsequent discussion mainly focuses on the recent algorithms for the influence maximization problem in social network diffusion and their applications.**

Index terms - Social Network, Influence node selection, information diffusion, influence maximization

I. INTRODUCTION

The rise of the Internet and the World Wide Web has enabled us to investigate large scale social networks, and there has been growing interest in social network analysis. Nowadays, the development of Internet has changed radically the way we communicate with each other. Communication helps us to better share knowledge, ideas and beliefs, thus influencing people behaviors.

In the recent decades, the rapid growth of Online Social Networks such as Facebook, Twitter and Google+ provide a nice platform for information diffusion and fast information exchange among their users.

Diffusion, according to Roger's definition [14], is the process by which an innovation is communicated through certain channels over time among the members of a social system. Three important elements: individual member, mutual interactions and communication channels are introduced from this definition, which are set as the basis for future analytical framework.

Later on, various diffusion models have been proposed to study the communication of an attitude or emotional state among a number of people , in a vast area such as widespread adoption in viral marketing [2,15, 16], information propagation on blogs [17, 18] and infectious diseases transmissions in epidemiology [19, 20].

This problem was first proposed by Kempe, Kleinberg, and Tardos [3] .They are the first to formulate the problem as a discrete optimization problem. The influence maximization problem is formally described as follows: given a social network represented by a(n) directed/undirected graph with nodes as users, edges are corresponding to social ties, edge weights are capturing influence probabilities, and a budget τ, which is a integer; the goal is to find a seed set of τ users such that by targeting these, the expected influence spread is maximized.

This paper surveys the recent advances in theoretical propagation models or diffusion models of online social networks, as well as the algorithms for the Influence Maximization problem. In section 3, an overview of existing diffusion models are presented. And with this framework, the next section 4, in surveys the various approaches for the influence maximization problem with high scalability. The last section, concludes the paper with some applications of social influence and information diffusion.

II. SOCIAL INFLUENCE AND INFLUENCE MAXIMIZATION

Social influence, as defined by Rashotte [22], is the change in an individual's thoughts, feelings, attitudes, and behaviors that results from interaction with other people or group. Social influence takes many forms and can been seen everywhere in Online Social Networks (OSNs). In the field of data mining and big data analysis, many applications such as viral marketing, recommendation systems and information diffusion are involved with social influence.

Influence maximization (IM) is one of the fundamental problems in studying social influence. For the reason that people are likely to be affected by decisions of their friends and colleagues, some researchers and marketers have investigated into social influence and the word-of-mouth effect in promoting new products and making profitable marketing strategies. Henceforth, the influence maximization problem has arisen: which key individuals should we target as the promising seeds in order to maximize the spread of influence?

In [1, 2], the influence maximization problem was studied in a probabilistic model of interaction, wherein selection of the most influential seeds were based on individual's overall effect on the network. In other works [3, 22, 4, 5, 23], many researchers take this seeding selection as a problem in discrete optimization. Formally, the influence maximization problem is defined as follows:

Influence Maximization [1]: Given a budget τ and a social network, which is represented as a directed graph G = (V, E), where users are represented as nodes and edges indicate their relationships, the goal is to select a seed set of τ users such that by initially targeting them, the expected influence spread (in terms of expected number of adopted users) can be maximized.

III. INFORMATION DIFFUSION MODELS

Influence diffusion is the process by which information propagates through certain intermediaries over time among the individuals of a social network. Currently, there are a variety of diffusion models arising from the economics and sociology communities. The most popular models are Linear Threshold model and Independent Cascading model, which are widely used in studying the social influence problems.
Besides these two well-known models, there are many variations and extensions of the models to reflect more complicated real-world situations. In this section, the recent literature on theoretical models of influence diffusion has been surveyed. The various models are depicted in figure 3.1

3.1 Threshold Models: In mathematical or statistical modelling, a threshold model is any model where a threshold value, or set of threshold values are used to distinguish ranges of values where the behavior predicted by the model varies in some important way. The threshold models were first proposed by Mark Granovetter [24] to model collective behavior, which aimed at treating binary decisions problems, such as diffusion of innovations, spreading rumors and diseases, voting and so on. He used the threshold model to explain the riot, residential segregation, and the spiral of silence.

In other words, this threshold represents the number of other agents in the population or local neighborhood following that particular activity. Each agent has a threshold that, when exceeded, leads the agent to
adopt an activity. In his model, each edge (v,u) is associated with a weight $w_{v,u}$, and each node v has a threshold $_v$ such that if the fraction of v's neighbors which are active exceeds v's threshold, then v will become active. Threshold models are especially useful in the structural analysis of collective action.

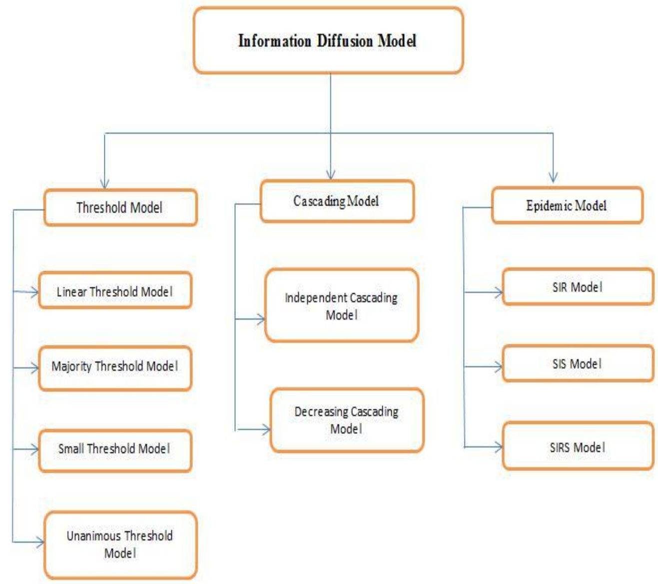

Figure 3.1: An overview of information diffusion models

3.1.1 Linear Threshold Model:

Linear Threshold (LT) model is the one that has been extensively used in studying diffusion models among the generalizations of threshold models.

In this model, each node v ε V satisfies the following inequality:

$$\sum_{u \rightarrow v \, \epsilon E} p(u,v) \leq 1 .$$

where p(u,v) is the probability of selecting the particular node for the diffusion of information. At the beginning, each node v chooses a threshold value $_v$ ϵ [0,1] uniformly and independently at random. An inactive node v becomes active in round i if

$$\sum_{u \rightarrow v \, \epsilon E} p(u,v) \leq {}_{v}.$$

The threshold in this model is related to a linear constraint of edge weight, and hence gets the name for the model. It is important to note that given the thresholds in advance, the diffusion process is deterministic, but we can still inject the randomness by randomizing the individual threshold.

Given the influence function μ(.), Kempe et al. [3] proved the following two theorems:

Theorem 1 For an arbitrary instance of the Linear Threshold Model, the objective influence function μ(.) is submodular.

Theorem 2 The influence maximization problem is NP-hard under the Linear Threshold model.

3.1.2 The Majority Threshold Model:

The Majority Threshold (MT) model is one of the most important and well-studied model, in which each vertex v ε V becomes active if the majority of its neighbours are active, that is the threshold $_v$ = 1/ 2 d(v). This model has many applications in voting systems, distributing computing and so on [25, 26]. Chen [27] shows that with the majority thresholds setting, the influence maximization problem shares the same hardness of approximation ratio as the general one.

3.1.3 The Small Threshold Model:

The other interesting case is the Small Threshold (ST) model, in which all thresholds are some small constants [29].
Intuitively, when the threshold $_v$ = 1, the influence maximization problem can be easily solved by selecting an arbitrary node in each connected component. However, Chen [27] shows that the hardness of approximation result continues to hold when each vertex's threshold $_v$ = 2. In addition, Dreyer [30] proves that if the threshold of any vertex is any $_v \geq 3$, the problem is NP-hard as well. $_v$ for

3.1.4 The Unanimous Threshold Model: In the Unanimous Threshold (UT) model, the threshold for each vertex is $_v$ = d(v), which is equal to its degree. With this setting, the UT model is the most influence resistant model among all the threshold models. This model is usually used in studying complex network security and vulnerability. For example, in an ideal virus resistant network, when the computer virus is spreading, a vertex can be affected if all of its neighbours have been infected. Based on this, the following theorem has been formulated and proved.

Theorem 3 If all thresholds in a graph are unanimous, the Influence Maximization problem is NP-hard.

3.2 Cascading Model: Inspired by the work on interacting particle systems [30, 31] and probability theory, dynamic cascade models are considered for the diffusion process. In the context of marketing, Goldenberg et al. [32,33] firstly studied the cascade models. In the cascade model, each active node u is only given one chance to activate each of its currently inactive neighbor v with the success probability p(u ,v), where u → v ϵ E.
If v has multiple active neighbors, their attempts to activate v are sequenced in an arbitrary order. In particular, if u is newly activated in round i, u →v ϵ E, v is inactive, and u is given a chance to activate v in round i+1, then v is activated by u with probability p(u, v). If u succeeds, then v becomes active at time i+1. If u fails to activate v, then u cannot activate v in subsequent rounds. The process continues until no more activation is possible.

3.2.1 Independent Cascading Model: To better describe the cascading models, one thing that needs to specified is that the probability for a newly activated node v to successfully make an attempt to activate its currently inactive neighbours u. The simplest case is Independent Cascading (IC) model, in which the probability is a constant $\rho_u(v)$, independent of the history of the diffusion process thus far.

In addition to that, to better define the model, one need to introduce the order-independence here. The order-independence indicates that the order of attempts made by each node in S does not affect the probability for u to be active in the end.

3.2.2 Decreasing Cascading Model:

Compared with the IC model, the Decreasing Cascading (DC) model [23] is more general and practical. The DC model naturally incorporates a restriction that the function $\rho_u(v|S)$ is non-decreasing in S, which indicates that $\rho_u(v|S) \leq \rho_u(v|T)$, where Sand T are set of nodes. This better reflects the information saturation problem in the real-world which means, the probability of a successful activation of a node u decreases if more people have already made the attempts. The DC model contains the IC model as a special case.

3.3 Epidemic Model: The epidemic has had a major impact on the life and politics of the country. Modeling

the infectious diseases became a matter of general interest in the 19th century. An epidemic model describes the transmission of contagious disease through individuals. In the recent century, it has been widely used to model computer virus infections and information propagations such as news and rumors.

3.3.1 SIR Model: The SIR (Susceptible-Infectious-Recovered) model first proposed by Kermack and

McKendrick[34]. In this model, it considers a fixed population which is divided into three distinct classes: Susceptible (S), Infectious (I), and Recovered (R). The individual goes through three consecutive states:

$$\mathbf{S \rightarrow I \rightarrow R}$$

And the dynamics of the model cascades in such a way: given a fixed population at a particular time t, there exists three groups of people, S(t) represents the number of people who are susceptible to the contagion, I(t) represents the number of people who have been infected and are capable of infecting those who are susceptible; R(t) is the number of people who have been infected and recovered, which means they are immune to be infected again in the future.

3.3.2 SIS Model: The SIS model consider a fixed population with only two compartments Susceptible S(t) and infected I(t), thus the flow of this model may be considered as follows:

$$\mathbf{S \rightarrow I \rightarrow S}$$

The SIS can be easily derived from the SIR model by simply considering that the individuals recover with no immunity to the disease, that is, individuals are immediately susceptible once they have recovered.

3.3.3 SIRS Model: The SIRS model is an extension of the SIR model. An individual can go through consecutive states:

S→I→R→S

The difference between this model and the SIR model is that, it allows the individuals of recovered group to leave and rejoin the susceptible group.

IV. Influence Maximization Algorithms

Having discussed the basic models of diffusion in social network, this section, presents the problem of influence maximization which is the task of finding the set of most influential nodes in the network. Several researchers have come up with various algorithm which are discussed here.

The influence function f (.) is submodular and monotone increasing. Exploiting these properties, Kempe et al. [3] present a simple greedy algorithm that approximates the problem with the ratio of 1-1/e-ε for any ε > 0. However, the worst-case running time of the naive greedy algorithm is $O(n^2(m+n))$, which is prohibitive for large-scale networks. Thus, considerable work has been done to improve it.

The algorithm that are used for identifying influence maximization nodes are as follows:

In this section, the recent algorithmic study such as CELF [4], CELF++ [23], MPINS [7] and LDAG [5] algorithms, which can obtain high scalability for influence maximization problems has been demonstrated.

4.1 Greedy Algorithm: It simply executes in k rounds, and in each round a new entry that gives the largest marginal gain in f will be selected.

```
Greedy_InfluenceMaximization( )
  {
    Input: Graph, integer k, any set function f
    Output: Seed set S
     Initialize S with empty set
     While (|S|≤k)
        {
       Select node which have maximum function weight;
              Update S with S ∪ {u};
       }
   Return S;
}
```

The worst case running time of this naive greedy algorithm is $O(n^2(m+n))$, which is prohibitive for large-scale networks.

4.2 CELF (Cost-Effective Lazy Forward selection): The most notable work is [4], where submodularity is exploited to develop an efficient algorithm called Cost-Effective Lazy Forward (CELF) selection algorithm, based on a lazy-forward optimization in selecting seeds. The idea is that marginal gain of a node in the current iteration cannot be better than its marginal gain in the previous iterations.

```
Greedy Algorithm optimized with CELF( )
    {
        Input: Graph, integer k, σm      // where σm  is marginal
gain.
        Output: Seed set S
        Initialize S and Q with empty set
        For each u ε V do
              {
                 Calculate marginal gain for each {u};
                 Initialize round with zero; Add u to
                 Heap Q;
           }

 While (|S|≤k)
  {
                      u ← root element in Q
          ; if (round == |S|)
        {Update S with S ∪ {u};
          Update Q with Q-{u};
        }
      Else
          {
             Update Marginal gain= σm(S ∪ {u}) - σm (S);
             Round = |S|;
             Reinsert u into Q and heapify.
}
     }
 Return S;
}
```

International Journal of Advanced Information Science and Technology (IJAIST) ISSN: 2319:2682
Vol.3, No.11, November 2014 DOI:10.15693/ijaist/2014.v3i11.19-26

CELF maintains a table < u,$\partial_u(S)$ > sorted on $\partial_u(S)$ in decreasing order, where S is the current seed set and

$\partial_u(S)$ is the marginal gain of u with respect to S. $\partial_u(S)$ is re-evaluated only for the top node at each step and the table is resorted when only it is necessary. If a node remains at the top, it will be picked as the next seed. In real implementation, a heap Q is employed to represent the priority of each node and maintain the sorted table information.

It is easy to see that this optimization avoids the re-computation of marginal gains of all the nodes in any iteration, except the first one. Therefore, from the experimental results, the CELF optimization leads to a 700 times speedup in the greedy algorithm shown in [3].

4.3 Degree Discount Algorithm: Even with the improved greedy algorithms, their running time is still large and may not be suitable for large social network graphs. Degree is frequently used for selecting seeds in influence maximization. Experimental results in [6] showed that selecting vertices with maximum degrees as seeds results in larger influence spread than other heuristics, but is still not as large as the influence spread produced by the greedy algorithms.

The general idea is as follows: let *v* be a neighbor of vertex *u*. If *u* has been selected as a seed, then when considering the selection of *v* as a new seed based on its degree, we should not count the edge *vu* towards its degree. Thus we should discount *v*'s degree by one due to the presence of *u* in the seed set, and we do the same discount on *v*'s degree for every neighbor of *v* that is already in the seed set. This is a basic degree discount heuristic applicable to all cascade models, and is referred to as SingleDiscount in [6]

DegreeDiscount(*G; k*)

{

initialize S with empty set;

for each vertex v do

{

Compute the degree of every vertex d_v;

$D(d_v) = d_v$;

initialize tv =0;

}

for I = 1 to k do

{

Select u = maximum {$D(d_v)$ };

Update S with S ∪ {u};

}

for each neighbor v of u and v ϵ V \ S do

{

$t_v = t_v + 1$;

update $D(d_v)$ with degree discount;

}

}

Return S;

}

Algorithm [4.3] implements the degree discount heuristic. Using Fibonacci heap, the running time of Algorithm [4.3] is O(k log n + m). Therefore, we can say that Degree Discount Heuristic is much faster than the original greedy algorithm.

4.4: MPINS(Minimum sized Positive Influential Node Set) Algorithm: Most of the existing work ignore negative influences among individuals or groups. Motivated by alleviating social problems, such as drinking, smoking and gambling, MPINS[7], take both positive and negative influences into consideration and propose a new optimization problem, named the Minimum sized Positive Influential Node Set (MPINS) selection problem, to identify the minimum set of influential nodes, such that every node in the network can be positively influenced by these selected nodes no less than a threshold α.

The objective of the MPINS selection problem is to identify a subset of influential nodes as the initial nodes.

such that, all the other nodes in a social network can be positively influenced by these nodes no less than a threshold α.

In [7], they propose a new heuristic in two phase: First, With the help of breath first search, they find the Maximal independent set(MIS) and Secondly, they employ the pre-selected MIS, denoted by *M*, as the initial active node set to perform the greedy algorithm MPINS-GREED

MPINS-GREEDY(G, threshold α)

{

Initialize I with set of Maximal Independent Set (M); // where I is initial seed set

while f(I) < $|V|\alpha$ do

{

choose $u \in V \setminus I$ to maximize f(I∪{u}) update I=I ∪ {u};

}

return I;

}

The proposed MPINS method selects a Maximal Independent Set (MIS) and with the help of this heuristic the number of iterations of the algorithm decreases drastically. For selecting an influential MIS, only a small number of iterations of MPINS-GREEDY are needed to find a solution for MPINS. However, the number of iterations for the greedy algorithm

International Journal of Advanced Information Science and Technology (IJAIST) ISSN: 2319:2682
Vol.3, No.11, November 2014 DOI:10.15693/ijaist/2014.v3i11.19-26

proposed for solving this type of problem is considerable larger compared to the number of iterations of MPINS-GREEDY.

4.5 ACO (Ant Colony Optimization) Influence Maximization Algorithm:

Influence Maximization problem has also been solved using swarm intelligence technique. In paper [8],ACO has been used to tackle the problem when there is a complex and broad family of diffusion models and properties of monotonicity and submodularity does not hold. In this case the greedy approach cannot be used. Therefore, swarm intelligence, specifically the Ant Colony optimization has been used to address the influence maximization problem.

It has been proven that the weight-proportional competitive liner threshold model does not have the properties of monotonicity and submodularity [9].

The algorithm first initializes all of the pheromone values according to the InitializePheromoneValue() function. An iterative process then starts, with the GenerateSolution() functionbeing used by all ants to probabilistically construct solutions to the problem based on a given pheromone model in each iteration. The EvaluateSolution() function is used to evaluate the quality of the constructed solutions and some of the solutions are used by the UpdatePheromoneValue() function to update the pheromone before the next iteration starts.

Algorithm 6: ACO_InfluenceMaximization()

```
{
      InitializePheromoneValue();
  While (termination conditions not met)
      {
          GenerateSolution();
          EvaluateSolution();
          UpdatePheromoneValue();
          }
    Return best solution;
}
```

The InitializePheromoneValue( ) function is used to initialize the pheromone values of all nodes of the constructed complete digraph. Initially, each node has a very small pheromone value of $\varepsilon \neq 0$. A possible solution is then created for each node by assembling the solution components as follows. Starting node i is added first, and each of its first-level neighbors are independently selected with probability p; then its second-level neighbors are selected, and so on, until k nodes are assembled in the solution.

The influence of the solution-which corresponds to the expected number of the adopters at the end of the diffusion process is then evaluated. The influences of the top-m solutions are then used as the pheromone laid down on all component nodes of the solution. Different solutions may lay down pheromone values on the same nodes, in which case all pheromone values of the same node are summarized.

Then, in the iterative process, all ants probabilistically construct solutions to the problem. In the GenerateSolution() function, each artificial ant generates a complete target set by choosing the nodes according to a probabilistic state-transition rule.

The EvaluateSolution() function is then used to evaluate the performance of each solution. To evaluate the performance of a solution the expected spread of the solution needs to be computed. Specifically, given a particular diffusion model, simulate the process 1000 times, and compute the average number of influenced nodes for each solution.

Based on the survey of the various diffusion models and influence maximization algorithms, Table-1 presents the application of each and the reference papers that deal with the problem.

Table 1: Models Listing and Comparisons

Name	Application	Property	Reference
Linear Threshold (LT)	Collective behavior, spreading rumors and diseases	The objective μ (.) is submodular, and IM is NP-hard	[3]
Majority Threshold (MT)	Voting system, distributed computing	IM is NP-hard	[25,26,27]
Small Threshold (ST)		$\Delta_V = 1$, select an arbitrary node in each connected component; $\Delta_V \geq 2$, IM is NP-hard	[27,28,29]
Unanimous Threshold (UT)	Network security and vulnerability	IM is NP-hard, 2-approximation algorithm	[2]
Independent Cascading (IC)	Collective behavior, promote new products	The objective σ(.) is submodular, and IM is NP-hard	[3]
Decreasing Cascading (DC)	Collective behavior, spreading information	IC is a special case of DC, and the objective σ(.) is submodular, and IM is NP-hard	[22]
Susceptible-Infectious-Recovered (SIR)	Transmission of contagious disease	$dS/dt = -\beta SI$, $dI/dt = \beta SI - \gamma I$, $dR/dt = \gamma I$	[34]

V. CONCLUSION

As social networking becomes more prevalent in the activities of millions of people on a day-to-day basis, both research study and practical applications on social influence will continue to grow. Furthermore, the size of the networks on which the underlying applications need to be used also continues to grow over time. Therefore, effective and efficient social influence methods are in high demand. This paper has presented an exhaustive survey of influence diffusion models and algorithms for Influence maximization in social network.

In the future, an important and challenging research area is to develop efficient, effective and quantifiable social influence mechanisms to enable various applications in social networks and social media. This area lies in the intersection of computer science, sociology, and physics. In particular, scalable and parallel data mining algorithms, and scalable database and web technology have been changing how sociologists approach this problem. Instead of building conceptual models and conducting small scale simulations and user studies, more and more people now rely on large-scale data mining algorithms to analyse social network data.

We also need to think beyond submodularity and monotonicity and try to implement this problem using Genetic algorithm and Nature inspired algorithm. The area is still in its infancy, and we anticipate that more techniques will be developed for this problem in the future.

References

[1] P. Domingos and M. Richardson. Mining the network value of customers. In Proceedings of the 7th ACM ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pages 57–66, 2001.

[2]M. Richardson and P. Domingos. Mining knowledge- Sharing sites for viral marketing. In *Proceedings of the 8th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 61–70, 2002

[3] D. Kempe, J. M. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In Proceedings of the 9th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pages 137–146, 2003.

[4] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. S. Glance. Cost-effective outbreak detection in networks. In Proceedings of the 13th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pages 420–429, 2007.

[5] W. Chen, Y. Yuan, L. Zhang, Scalable influence maximization in social networks under the linear threshold model, in ICDM, Sydney, 2010.

[6] M. Kimura and K. Saito. Tractable models for information diffusion in social networks. In *Proceedings of the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases*, pages 259–271, 2006.

[7] Jing He ,Shouling Ji, Xiaojing Liao,Raheem Beyah. Minimum sized Positive Influential Node Set Selection for Social Network: Considering Both Positive and Negative Influences. IEEE 2013 Transaction Number 978-1-4799-3214-6/13.

[8] Application of the Ant Colony Optimization Algorithm to Competitive Viral Marketing Wan-Shiou Yang and Shi-Xin Weng Department of Information Management, National Changhua University of Education,No. 1, Jin-De Rd., Changhua 50007, Taiwan

[9] Borodin, A., Filmus, Y., Oren, J.: Threshold Models for Competitive Influence in Social Networks. In: Saberi, A. (ed.) WINE 2010. LNCS, vol. 6484, pp. 539–550. Springer, Heidelberg (2010)

[10] Opsahl, T., Agneessens, F., Skvoretz, J.: Node Centrality in Weighted Networks: Generalizing Degree and Shortest Paths. Social Networks 32(245), 245–251 (2010)

[11] F. Provost, B. Dalessandro, R. Hook, X. Zhang, and A. Murray. Audience selection for on-line brand advertising: privacy-friendly social network targeting. In *KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 707–716, New York, NY, USA, 2009. ACM.

[12] A. Goyal, F. Bonchi, and L. V. Lakshmanan. Discovering leaders from community actions. In *Proceeding of the 17th ACM conference on Information and knowledge management (CIKM'08)*, pages 499–508, 2008.

[13] N. Agarwal, H. Liu, L. Tang, and P. S. Yu. Identifying the influential bloggers in a community. In *Proceedings of the 1st ACM International Conference on Web Search and Data Mining (WSDM'08)*, pages 207–217, 2008.

[14] E. M. Rogers. Diffusion of innovations. New York: Free Press, 1962.

[15] P. Domingos. Mining social networks for viral marketing. In IEEE Intelligent Systems, 2005.

[16] J. Leskovec, A. Singh, and J. Kleinberg. Patterns of influence in a recommendation network. In In Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), 2006.

[17] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. In Proceedings of the 7th ACM Conference on Electronic Commerce,EC '06, pages 228–237. ACM, 2006.

[18] J. Leskovec, M. McGlohon, C. Faloutsos, N. Glance, and M. Hurst. Cascading behavior in large blog graphs. In In SIAM International Conference on Data Mining (SDM), 2006.

[19] P. S. Dodds and D. J. Watts. Universal behavior in a generalized model of contagion. Physical Review Letters, 92, 2004.

[20] N. Berger, C. Borgs, J. T. Chayes, and A. Saberi. On the spread of viruses on the internet. In In Proceedings of the 16th ACM-SIAM Symposium on Discrete Algorithm (SODA), 2005.

[21] L. Rashotte. Social influence. In the blackwell encyclopedia of sociology, IX:4426–4429, 2006.

[22]. D. Kempe, J. Kleinberg, and E. Tardos. Influential nodes in a diffusion model for social networks. In 32nd international conference on Automata, Languages and Programming, ICALP05, pages 137–146, Berlin, Heidelberg, 2005.Springer-Verlag.

[23] Y. Wang, G. Cong, G. Song, and K. Xie. Community-based greedy algorithm for mining top-k influential nodes in mobile social networks. In 16th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD'10, pages 1039–1048, New York, NY, USA, 2010.

[24] M. S. Granovetter. Threshold models of collective behavior. The American Journal of Sociology, 83(6):1420–1443, 1978.

[25] D. Peleg. Local majority voting, small coalitions and controlling monopolies in graphs: A review. In in Proceedings of the 3rd Colloquium on Structural Information and Communication Complexity, pages 170–179, 1996.

[26] D. Peleg. Size bounds for dynamic monopolies. Discrete Applied Mathematics, 86:263–273, 1998.

[27] C. Ning. On the approximability of influence in social networks. In Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms, 2008.

[28] F. S. Roberts. Graph-theoretical problems arising from defending against bioterrorism and controllingthe spread of fires. In

DIMACS/DIMATIA/Renyi Combinatorial Challenges Conference, 2006.

[29] P. A. Dreyer. Applications and variations of domination in graphs. Ph.D. Thesis, Rutgers University, 2000.

[30] R. Durrett. Lecture notes on particle systems and percolation. Wadsworth Publishing, 1988.

[31] T. M. Liggett. Interacting particle systems. Springer, 1985.

[32]J. Goldenberg, B. Libai, and E. Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. Marketing Letters, 3(211- 223), 2001.

[33] J. Goldenberg, B. Libai, and E. Muller. Using complex systems analysis to advance marketing theory develop ent. Academy of Marketing Science Review,2001.

[34] M. Kermack. Contributions to the mathematical theory of epidemics. In RoyalSociety of Edinburgh. Section A. Mathematics, volume 115, 1972.

Authors Profile

Lal Vikram Singh received the **B.TECH** degree in computer Science and engineering from Institution of Electronic and Telecommunication (IETE) in 2013.Currently doing **M.TECH** in computer science and engineering in Pondicherry University, Puducherry, India. His research interest includes Social Network Optimization, Cloud Computing, Big Data, Data Mining and Genetic Algorithm.

Dr. S Sivasathya received the **Ph.D** in Computer Science from Pondicherry University in 2009.Currently she is **Associate Professor** in Department of Computer Science, Pondicherry University, Puducherry, India. Her area of research and specialization includes Genetic Algorithm, Bioinformatics, Intrusion Detection and Response and Computer Based Music Analysis.