Targeted Influence Maximization through a Social Network

Rama Kumar Pasumarthi
Big Data Labs, American Express, India.
ramakumar.pasumarthi@aexp.com

Ramasuri Narayanam
IBM Research, India
ramasurn@in.ibm.com

Balaraman Ravindran
Department of CSE, IIT Madras
Chennai, India
ravi@cse.iitm.ac.in

Abstract

In this paper, we address the problem of Targeted Influence Maximization (TIM) through a social network. Often companies want to promote their products to certain type of customers as opposed to targeting the entire social network. That is, there is a need to maximize influence over a targeted audience in the network. Towards this end, we present a novel objective function for the targeted influence maximization problem. It turns out that this objective function is the difference between two relevant submodular functions. By building upon the recently developed theory for optimizing the difference between two submodular functions, we develop an efficient algorithm to maximize our objective function. In detailed experiments on both synthetic and real world data sets, we show that the quality of solution for TIM improves using our proposed approach.

1 Introduction

It is well known that social networks play a crucial role in the spread of information and influence among themselves [12, 36, 33, 20, 3]. It is important to study the dynamics of information diffusion in order to understand the extent to which such ideas are adopted over the network and this field has a long history in social sciences [33, 20, 34]. For instance, diffusion processes have been investigated for viral marketing [21, 25, 23, 9]. The phenomenon of viral marketing is to exploit the social interactions among individuals to promote awareness for new products. Domingos and Richardson [11], Richardson and Domingos [31] posed a fundamental algorithmic problem in the context of viral marketing with single product as follows. We are given the information about the extent individuals influence each other. We would like to market a new product that we hope will be adopted by a large fraction of individuals in the network. One of the key issues in viral marketing is to select a set of influential individuals (also called as *initial seeds*) in the social network and give them free samples of the product (or simply promotional offers on the product) to trigger cascade of influence over the network. The problem is, given a (integer) value for k, how should we choose a set of k influential individuals so that the cascade of influence over the network is maximized? Hereafter, we refer to this problem as *influence maximization problem* [21]. It is shown to be a NP-hard problem [11, 31, 21, 25] and we refer to Section 2 for more details on the related work.

Most of the relevant literature on viral marketing of products focus on the entire social network to maximize the influence. However, companies often want to promote their products to certain type of customers as opposed to targeting the entire social network. For instance, the company introducing a new four-wheeler insurance plan should target only those having a four-wheeler, but not those having a two-wheeler. That is, it is important to consider the problem of maximizing the influence over a targeted set of individuals in a social network as opposed to the entire network. Towards this end, in this paper, we introduce the problem of targeted influence maximization (TIM) through a social network and present a new framework that address this problem.

Our Approach: In the given social network, consider a portion of individuals as the target set, call it A, within which we want to maximize the influence. Given a (small integer) value for k, we want to find k seeds so that the influence spread within A is maximum. We call a node *active* if it is influenced by its neighbor(s). It is important to note that some of these k seeds need not belong to the set A. For instance, consider a stylized social network as shown in Figure 1. Here assume that the set of nodes filled with black color represent the target set A. Now, we want to make the following two intuitive observations:

- If we were to choose a single node as a seed, then the solution could possibly be node A which does not belong to the target set.

- Furthermore, assume that node the target node E is already influenced. Node E may influence the non-target node F, which can further influence the target node C. Here the non-target node F acts as an intermediate node to form a path between two target nodes.

That is, in the process of influence maximization within A, from this example, it is clear that *we may approach the nodes* not in A *either as initial seeds or as intermediate nodes to reach out the targeted nodes.*

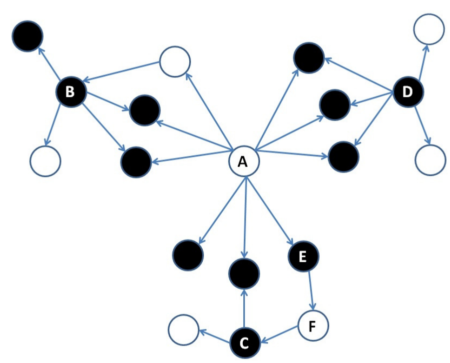

Figure 1: A stylized example of social network

Role of Incentives in Viral Marketing: Also, it is known that diffusion of influence over social networks (i.e. viral marketing) works best when valuable and tangible incentives are offered [35]. For example, an athletic clothing multichannel retailer recently offered an incentive when it launched a viral marketing campaign that rewarded message recipients with a free T-shirt and a $1 donation to a charitable trust when an individual sent the special email message to five friends and three of those friends opted in to the retailer's catalog [5]. This campaign was tremendously successful and it caused the click-through rate three times higher than normal.

Now consider a company interested in the targeted advertising of its product. Since the companies often have budget constraints, the company would want to minimize the incentives offered to non-targeted nodes while maximizing the influence in the targeted set A. Thus, a reasonable objective function for the targeted influence maximization problem would be to maximize the difference between the expected number of active target nodes in A and the expected number of active non-target nodes. It turns out that this objective function is the difference between two submodular functions. We then present an efficient algorithm to maximize our objective function by building upon the recently developed theory for optimizing the difference between two submodular functions [32]. We also present a few scalable heuristics for the targeted influence maximization problem.

1.1 Outline of the Paper

We organize the as follows. We briefly review the relevant literature in Section 2. We present the model and algorithms for the targeted influence maximization in Section 3. We conduct thorough experimental evaluation of our approach in Section 4. We conclude the paper with a few pointers to future work in Section 5.

2 Relevant Work

We first briefly mention the models for diffusion of information. There are two well known operational models in the literature that capture the underlying dynamics of the information propagation in viral marketing. They are the linear threshold model [34, 18, 21] and the independent cascade model [17, 21].

Domingos and Richardson [11, 31] were the first to study influence maximization problem as an algorithmic problem. They modeled social networks as Markov random fields where the probability of an individual adopting a technology (or buying a product) is a function of both the intrinsic value of the technology (or the product) to the individual and the influence of neighbors. The computational aspects of the influence maximization problem are investigated by Kempe, Kleinberg, and Tardos [21]. The authors show that the optimization problem of selecting the most influential nodes is NP-hard and derive the first provable approximation guarantees for the proposed algorithm. Recall that the objective function, $\sigma(.)$, for information diffusion is the expected number of nodes (i.e. $\sigma(S)$) that become active at the end of the diffusion process for a given set of initial active nodes (i.e. S). The authors first show that this objective function is a sub-modular function under both the linear threshold model and the independent cascade model. A function $g(.)$ is called sub-modular if it satisfies $g(S \cup \{i\}) - g(S) \geq g(T \cup \{i\}) - g(T)$ for all elements i and all pairs of sets $S \subseteq T \subseteq N$ where N is the set of nodes in the graph.

The authors [21] then propose a greedy approximation algorithm for the influence maximization problem and they showed that this greedy algorithm achieves an approximation guarantee of $(1 - \frac{1}{e})$ where $e = \sum_{r=1}^{\inf} \frac{1}{r!}$. We note that the running time of this greedy approximation algorithm [21] is $O(knRm)$ as mentioned in Chen, Wang, and Yang [9]. Here R is the number of repetitions of the each experiment, n is the number of nodes, and m is the number of edges in the graph. Leskovec, et. al. [25] proposed an efficient algorithm for the influence maximization problem based on the submodularity of the underlying objective function that scales to large problems and is reportedly 700 times faster than the greedy algorithm of Kempe, Kleinberg, and Tardos [21]. Chen, Wang, and Yang [9] present an efficient algorithm to find the initial seeds in a social network and this algorithm improves upon the greedy algorithm of Kempe, Kleinberg, and Tardos [21] and also the algorithm of Leskovec, Krause, and Guestrin [25] in terms of its running time. Even-Dar and Shapira [14] study the influence maximization problem in the context of probabilistic voter model. Kimura and Saito [22] proposed a shortest-path based influence cascade model and designed efficient algorithms for finding the most influential nodes. Ramasuri [29] and Chen *et al.* [7] are a few advances in the literature in the context of the influence maximization problem. We note that the ideas similar to those in the context of influence maximization problem are utilized to design immunization strategies in the context of virus propagation [16].

3 Proposed Approach to Targeted Influence Maximization

In this section, we first present the model for the TIM problem and then we describe our solution approach to the TIM problem.

3.1 The Model for Diffusion of Influence

We use the continuous time *Independent Cascade* model as the generative process for diffusion of influence over social networks [37]. In this model, each pair of users, u and v, have an *influence strength* associated with them, and a *transmission function*, $\delta_{uv}(\tau)$, where τ, the *incubation time*, corresponds to the time it takes for v to get influenced once u is influenced. The pairwise influence strengths can be seen as parameters of the transmission function, given a pair of users.

Independent Cascade Model: Given a directed network $G = (V, E)$, with set of nodes V and set of edges E, and pairwise transmission functions and corresponding influence strengths, the diffusion model begins with a set of infected source nodes, S, at time $t = 0$. These can be seen as the initial adopters of a certain idea or contagion. Then the diffusion mechanism progresses as follows: Whenever a node u gets infected, say at time t_u, for each of it's followers v, (direct neighbors of u), it draws a sample transmission time τ from δ_{uv} and proposes (independently) an infection time $t_{uv} := t_u + \tau$ for node v. Amongst all proposed times for v, v chooses the earliest timestamp as it's infection time. The diffusion progresses so on in a cascading manner.

We follow the SIR (*Susceptible-Infected-Recovered*) model for the state of nodes, where each node starts in a susceptible state, can get infected only once, and is in a recovered state after that. Note that each cascade (the spread of one particular contagion) is in the form of a Directed Acyclic Graph (DAG), even though the graph itself need not be a DAG.

Transmission functions: Distributions such as Weibull (which captures both Exponential and Rayleigh) and Power law have been used to model the transmission functions [37]. The parameters of Weibull or Power law correspond to pairwise influence strengths.

3.2 Supermodular-submodular (Sup-Sub) procedure for maximizing the difference between two submodular functions

Recall that A denotes the set of target nodes. Let S denote set of seed nodes. Let T correspond to the time window of observation. Let $\sigma_A(S, T)$ be the expected number of target nodes infected within time window T given that nodes in S were seed nodes.

We define the objective function for the problem of targeted influence over a social network as:

$$\arg \max_{S:|S|=k} \sigma_A(S, T) - \sigma_{V \setminus A}(S, T)$$

where

$$\begin{aligned} \sigma_A(S, T) &= E\left[\sum_{v \in A} I(t_v < T)\right] \\ &= \sum_{v \in A} Pr(t_v < T) \end{aligned}$$

where $I(t_v < T)$ takes value 1 if node v gets activated within T, otherwise 0. We now present the following two useful results.

Lemma 1 *$\sigma_A(., .)$ is a submodular function.*

Lemma 2 *$\sigma_{V \setminus A}(., .)$ is a submodular function.*

The above two lemmas can be proved using the arguments similar to that in Kempe, Kleinberg, Tardos [21]. We omit the proofs in the interest of space in this paper.

Henceforth, let $f(S)$ denote $\sigma_A(S, T)$ and $g(S)$ denote $\sigma_{V \setminus A}(S, T)$.

Let $v(S) := f(S) - g(S)$, which is the objective function we wish to maximize in Targeted Influence Maximization.

Iyer *et. al* [32] have outlined a procedure, called *Sup-Sub* procedure, for maximizing (or minimizing) the difference of two submodular functions. This is shown in Algorithm 1. In this algorithm, the term $m^g_X(Y)$ represents a modular upper bound for any submodular function (g). More formally, it is given by [32]:

$$m^g_X(Y) = g(X) - \sum_{v \in X \setminus Y} [g(X) - g(v)] + \sum_{v \in Y \setminus X} g(v) \tag{1}$$

Note that the inner loop consists of submodular function maximization, where we can use the well known greedy algorithm to optimize it.

Algorithm 1: Sup-Sub procedure

Objective: Maximise $f(X) - g(X)$
$X^0 = \phi, t = 0$
while $X^t \neq X^{t+1}$ **do**
  $X^{t+1} :=$ arg max$_Y f(Y) - m^g_{X^t}(Y)$
  $t = t + 1$

Let $h_X(Y) := f(Y) - m^g_X(Y)$ denote the submodular function in the inner loop, which we wish to maximize in a given iteration. We implemented the traditional greedy approach, with lazy evaluation (**CELF**). [40] As a part of this computation, we need to compute the marginal gain on adding a node u to existing set Y.

The marginal gain over function $h_X(Y)$ for adding an additional node u to existing set Y is

$$\begin{aligned} h_X(u|Y) &:= f(Y \cup \{u\}) - f(Y) - [m^g_X(Y \cup \{u\}) - m^g_X(Y)] \\ &= g(Y \cup \{u\}) - f(Y) - g(X) + g(X - \{u\}) \text{ if } u \in X \\ &= f(Y \cup \{u\}) - f(Y) - g(\{u\}) \text{ if } u \notin X \end{aligned}$$

Since we cannot say whether $h_X(u|Y)$, the marginal gain considered above, is always non-negative, we make the following modification: We will add nodes using the greedy algorithm, until we add the required number of k nodes or the maringal gain of node being added, $h_X(u|Y)$, is non-positive.

The SupSub procedure for TIM, with the above modification is presented below.

Algorithm 2: Sup-Sub procedure for TIM

$S^0 = \phi, t = 0$
while $S^t \neq S^{t+1}$ **do**
  $S = \phi$
  while $|S| < k$ **do**
    $u =$ arg max$_u h_{S^t}(u|S)$
    if $h_{S^t}(u|S) > 0$ **then**
      $S = S \cup$ arg max$_u h_{S^t}(u|S)$
    else
      break
  $S^{t+1} = S$
  $t = t + 1$

Note that, in this algorithm, we need to compute the marginal influences of the nodes ($h()$). In this paper, we utilize Cohen's randomized algorithm [10] to efficiently compute these marginal influences of the nodes to solve the TIM problem.

4 Experimental Results

In this section, we empirically evaluate the performance of the proposed algorithm using synthetic as well as real-world data sets. We first outline a few heuristics for the TIM problem based on the benchmark methods available for the standard influence maximization problem [21]. We then describe the data sets and present the experimental results.

4.1 Heuristics

Here present three heuristics for the TIM problem.

(i) Targeted-set restricted Discounted Maximum Degree Heuristic (TD-MDH): The concept of centrality is well addressed in social networks (Freeman [79]). One of the simplest and best-known measures of centrality is degree centrality, which is a count of the number of edges incident upon a given node. Motivated by the degree discount heuristic [8], here we present a similar approach for the TIM problem.

For each node $u \in V$, we define its degree restricted to the target set A as the number its neighbors in A. That is,

$$d^A(u) = \{v \in A \mid (u, v) \in E\} \tag{2}$$

Let S_v be the set of neighbors of node $v \in V$ in the set S. Algorithm 3 outlines this method.

Algorithm 3: TD-MDH

Input: Graph $G = (V, E)$ and the objective function $\sigma^A(.)$
Output: A set S of k seed nodes

for all $u \in V$ **do** Compute $d^A(u)$;
Set $S \leftarrow \Phi$;
for $i = 1$ **to** k **do**
  Find node $v \in V \setminus S$ with maximum value for $d^A(v) - \big(2|S_v| + \big(d^A(v) - |S_v|\big)\,|S_v|p\big)$
  $S \leftarrow S \cup \{v\}$

(ii) Targeted-set restricted Maximum Degree Heuristic (T-MDH): Following this method, we choose the k seeds as those nodes from the target set A having high degree restricted to A. Algorithm 4 outlines this method.

Algorithm 4: T-MDH

Input: Graph $G = (V, E)$ and the objective function $\sigma^A(.)$
Output: A set S of k seed nodes

for all $u \in V$ **do** Compute $d^A(u)$;
Set $S \leftarrow \Phi$;
for $i = 1$ **to** k **do**
  Find node $v \in V \setminus S$ with maximum value for $d^A(v)$
  $S \leftarrow S \cup \{v\}$

(iii) Targeted-set restricted Random Heuristic (T-Random): Here we choose the k seed nodes uniformly at random from the set A.

4.2 Network Data Sets

Methodology: For each of the networks, we compare the propsoed method, SupSub for TIM with the two baselines, T-MDH and TD-MDH using the value of the TIM-objective function over k initial seed nodes. This is carried out for $k = 5, 10, 15, 20$ and 25. The transmission function is modeled as an exponential distribution. We set the time window for influence spread, T, to be 10.

Synthetic Networks: We used synthetic graphs of the type Core-Periphery (which have a strong core component to which several peripheral nodes connect). For each of the edges, the mean of the exponential distribution is sampled uniformly from $(0, 10)$. We generated networks with 1024 nodes, 2048 nodes and 4096 nodes.

Real World Networks: we used Netscience real world network, which is a co-authorship network of scientists working on network theory and experiment, compiled by M. Newman in 2006 [38]. It has 4204 nodes and 2742 edges. This is a weighted graph. The weights on the directed edges were treated as the mean of the exponential distribution for modelling the transmission function.

4.3 Results

In this section, we compare the performance of SupSup-TIM, with the two baselines; TD-MDH and T-MDH over the above mentioned synthetic and real world networks. Half of the nodes in the network were chosen as target nodes at random. From these graph plots, it is clear that our proposed solution approach to TIM problem clearly outperforms that of the standard benchmarks.

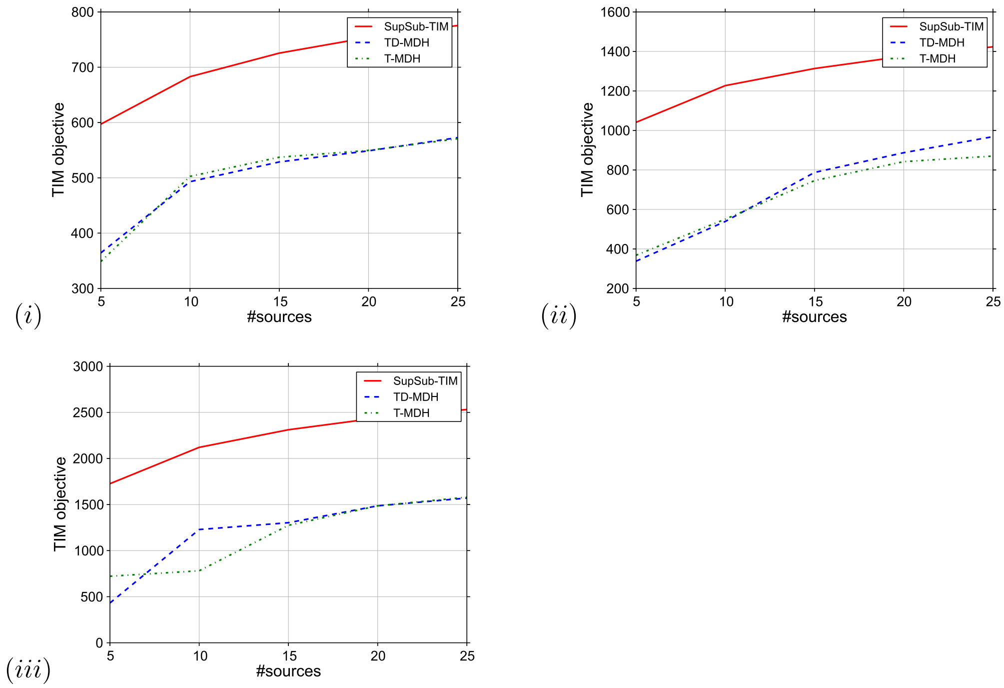

Figure 2: TIM objective on synthetic graphs for varying k using three synthetic graphs having (i) 1024 nodes, 2048 edges, (ii) 2048 nodes, 4096 edges and (iii) 4096 nodes, 8192 edges

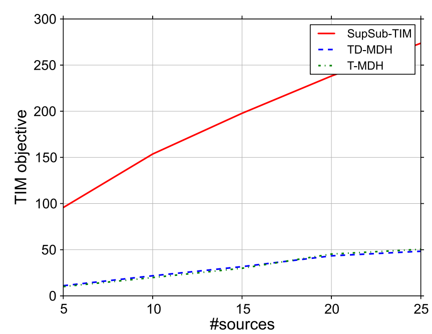

Figure 3: TIM objective on synthetic graphs for varying k using Netscience graph

5 Conclusions and Future Work

In this paper, we proposed a novel objective function for the problem of targeted influence maximization and this objective function turned out to be the difference between two submodular functions. We then presented an efficient algorithm based on the submodular function maximization framework [32]. We also outlined a few heuristics for the TIM problem based on the standard techniques in the relevant literature. There are several ways to further improve the work in this paper. One potential direction of this kind of work is to design scalable heuristics/algorithms for the TIM problem. Another potential direction is to derive approximation guarantees on the quality of the solution for the TIM problem since we work with heuristics for solving the underlying objective function.

References

[1] C. Aggarwal, A. Khan, and X. Yan. On flow authority discovery in social networks. In *Proceedings of SIAM Conference on Data Mining (SDM)*, pages 522–533, 2011.

[2] G. Calinescu, C. Chekuri, M. Pal, and J. Vondrak. Maximizing a submodular set function subject to a matroid constraint. In *Proceedings of IPCO*, pages 182–196, 2007.

[3] A. Calvo-Armengol. Job contact networks. *Journal of Economic Theory*, 115(1):191–206, 2004.

[4] C. Chekuri and A. Kumar. Maximum coverage problem with group budget constraints and applications. In Proceedings of APPROX, 72-83, 2004.

[5] http://www.clickz.com/clickz/column/1700226/tips-optimizing-viral-marketing-campaigns

[6] G. Cornuejols, M. Fisher, and G. Nemhauser. Location of bank accounts to optimize oat: an analytic study of exact and approximate algorithms. Management Science, 23:789-810, 1977.

[7] W. Chen, A. Collins, R. Cummings, T. Ke, Z. Liu, D. Rincon, X. Sun, Y. Wang, W. Wei, and Y. Yuan. Influence maximization in social networks when negative opinions may emerge and propagate. In *Proceedings of SIAM SDM*, 2011.

[8] W. Chen, S. Yang, and Y. Wang. Scalable Influence Maximization for Prevalent Viral Marketing in Large-Scale Social Networks. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2010.

[9] W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 937–944, 2009.

[10] E. Cohen. Size-estimation framework with applications to transitive closure and reachability. Journal of Computer and System Sciences, 55(3):441–453, 1997.

[11] P. Domingos and M. Richardson. Mining the network value of customers. In *Proceedings of the 7th SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 57–66, 2001.

[12] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press, Cambridge, UK, 2010.

[13] P. Estevez, P. Vera, and K. Saito. Selecting the most influential nodes in social networks. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, pages 2397–2402, 2007.

[14] E. Even-Dar and A. Shapira. A note on maximizing the spread of influence in social networks. In *Proceedings of the 3rd Workshop on Internet and Network Economics (WINE)*, pages 281–286, 2007.

[15] M.L. Fisher, G.L. Nemhauser and L.A. Wolsey. An analysis of approximations for maximizing submodular set functions - II. Mathematical Programming Study, 8:73-87, 1978.

[16] C. Gao, J. Liu, and N. Zhong. Network immunization and virus propagation in email networks: experimental evaluation and analysis. In *Knowledge and Information Systems*, 27(2):253-279, 2011.

[17] 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*, 12(3):211-223, 2001.

[18] M. Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, 83:1420–1443, 1978.

[19] X. He, G. Song, W. Chen, and Q. Jiang. Influence blocking maximization in social networks under the competitive linear threshold model. In Proceedings of the 12th SIAM International Conference on Data Mining (SDM), 2012.

[20] M. Jackson. *Social and Economic Networks*. Princeton University Press, Princeton and Oxford, 2008.

[21] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the 9th SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 137–146, 2003.

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

[23] J. Leskovec, L. Adamic, and B. Huberman. The dynamics of viral marketing. *ACM Transactions on the Web*, Volume 1, Number 1, Article Number 5, 2007, 39 pages. DOI = 10.1145/1232722.1232727 http://doi.acm.org/ 10.1145/1232722.1232727

[24] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Signed networks in social media. In *Proceedings of the 28th ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 1361-1370, 2010.

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

[26] A. Nanavati, R. Singh, D. Chakraborty, K. Dasgupta, S. Mukherjea, G. Das, S. Gurumurthy, and A. Joshi. Analyzing the structure and evolution of massive telecom graphs. *IEEE Transactions on Knowledge Discovery and Data Engineering*, 20(5):703–718, 2008.

[27] G. Nemhauser, L. Wolsey, and M. Fisher. An analysis of approximations for maximizing submodular set functions-1. *Mathematical Programming*, 14(1):265–294, 1978.

[28] M. E. J. Newman. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences (PNAS)*, 98:404-409, 2009.

[29] N. Ramasuri and Y. Narahari. A Shapley Value based Approach to Discover Influential Nodes in Social Networks. *IEEE Transactions on Automation Science and Engineering (IEEE TASE)*, 8(1):130-147, 2011.

[30] M. Richardson, R. Agrawal, and P. Domingos. Trust management for the semantic web. In *Proceedings of ISWC*, 2003.

[31] M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the 8th SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 61–70, 2002.

[32] R. Iyer and J.A. Bilmes. Algorithms for Approximate Minimization of the Difference between Submodular Functions. In Uncertainity in Artificial Intelligence (UAI), pages 407–417, 2012.

[33] E. Rogers. *Diffusion of Innovations*. Free Press, New York, USA, 1995.

[34] T. Schelling. *Micromotives and Macrobehavior*. W.W. Norton and Company, 1978.

[35] Y. Singer. *How to win friends and influence people, truthfully: influence maximization mechanisms for social networks*. In Proceedings of WSDM, pages 733–742, 2012.

[36] S. Wasserman and K. Faust. *Social Network Analysis*. Cambridge University Press, 1994.

[37] Manuel Gomez-Rodriguez, David Balduzzi, and Bernhard Scho lkopf. Uncovering the temporal dynamics of diffusion networks. In *ICML, pages 561568, 2011.*

[38] M. E. J. Newman Finding community structure in networks using the eigenvectors of matrices Preprint physics*/0605087 (2006).*

[39] Leskovec, J., Chakrabarti, D., Kleinberg, J., Faloutsos, C., and Ghahramani, Z. Kronecker graphs: An approach to modeling networks. The Journal of Machine Learning Research*, 11, 985-1042, 2010*

[40] Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., and Glance, N. Cost-effective outbreak detection in networks. *In* Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining *(pp. 420-429). ACM.. 2007*