

Multi-level Revenue Sharing for Viral Marketing

Zeinab Abbassi *
Columbia University
New York, NY

Vishal Misra †
Columbia University
New York, NY

Abstract

In this paper we present the design and analysis of revenue sharing schemes for viral marketing over social networks. The increasing need for monetizing social networks more effectively is causing social network platforms to look for alternatives to online behavioral targeting. Specifically, we turn to cooperative game theory and the Shapley value to design revenue sharing schemes to incentivize users to help the social network platform for more effective viral marketing. Our goal is to identify mechanisms that achieve desirable objectives in terms of computability, individual rationality, and potential reach. In particular, we propose multi-level revenue sharing for referral-based and viral marketing over online social networks. We show via simulations that users have more incentive to collaborate with the social network platform in implementing the campaign when the revenue or discount is shared across multiple levels rather than the commonly used single-level model. For this purpose, we design the graph-based model, for which we show that computing the Shapley value is $\#P$ -hard. However, we show that in a variation of that model, which we call the tree-based model, computing the Shapley value becomes polynomial time. We also show that the revenue function is supermodular only in the tree-based model. Supermodularity of the revenue function entails desirable corollaries.

1. INTRODUCTION

In recent years, the Web has been, among other things, leveraged to harness the power of users to carry out tasks that require collective efforts. Wikipedia, crowdsourcing platforms such as Amazon’s Mechanical Turk, and Yahoo! Answers are only a few examples. Users participate with different

*zeinab@cs.columbia.edu

†misra@cs.columbia.edu

This work was supported in part by the NSF grant CNS-1017934. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

incentives such as monetary compensation, social recognition, altruism or a combination of these. Some of these tasks require networked structures to succeed. The recent tremendous growth in the popularity of online social networks suggests leveraging these networks for those types of tasks. The recent DARPA network challenge is an example of such a task. The challenge was to locate ten red balloons spread all over the United States on a given date as quickly as possible for the prize of \$40000 [1]. A task that would be considered impossible using conventional information gathering methods [2]. The winning team implemented a recursive incentive scheme over a social network of referrals with the objective to locate the balloons in minimum time [3].

Refer-a-friend marketing (which is a type of viral marketing) is another example of such tasks because the network of trust between the referrer and the potential adopters plays an important role in the adoption of a product in a viral marketing campaign. In a refer-a-friend advertising campaign, typically, a current user gets some form of discount for referring a product to her friend that ends up in the adoption of the product. Referring a friend might trigger a cascade if the new adopter recommends the product to her friends as well (hence the term viral marketing).

Another motivating scenario inspired by the above ideas can be for advertising over online social networks. The increasing need for monetizing social networks more effectively is causing social network platforms to look for alternatives to online behavioral targeting. A specific example for this model is as follows (illustrated in Figure 1): A social network platform may build the following system to target ads and coupons more effectively. This system allows users to opt in and help the platform in exchange for a share in the revenue. To be more precise, a user who opts in this system is presented with a number of ads/coupons to assign to a limited number of her friends according to their interests. In this example a social network user, named Alice, is presented with two ads: one about audiobooks and the other about places to travel with a baby. Alice knows that Bob has a baby and might be interested in going on a trip, so she would select Bob for the ad: "Best places to travel with a baby". On the other hand, knowing that Carol enjoys listening to audio books, Alice would assign the ad for free audio books to her. This will help the social network route ads to the right users. Bob sees the ad about traveling with a baby as he logs into his account. He also thinks that another fellow parent might be interested in this ad, therefore,

he suggests it to be shown to her. The source of the ad can be transparent to the recipient of the ad. In other words, this system can be incorporated in the ad network that the social network platform implements.

A significant challenge for this model to work is to give the users the proper motivation to earnestly contribute to the system in finding the most relevant match. To incentivize users, the social network platform shares the added revenue with the users that opt in and have impact. In this paper, we introduce various revenue sharing models and discuss fairness and individual rationality of such incentive schemes, and design efficient algorithms to compute and implement such schemes. In particular, we propose models in which referrals (both for products and ads) are rewarded either for one level or multiple levels, and discuss Shapley value revenue sharing. In other words, we discuss single-level and multi-level propagation models, and identify tree-based multi-level propagation as a special case of the graph-based propagation model. We compare these models in terms of the polynomial-time computability of Shapley revenue shares, individual rationality of the revenue shares, and potential reach or expected effectiveness of these models. First, we prove that finding Shapley value revenue shares is $\#P$ -hard for general graph-based multi-level models, but it is polynomial-time solvable for single-level and tree-based multi-level propagation models. Further, by showing supermodularity of revenue function for tree-based propagation models, we show that for the tree-based propagation model, (i) Shapley value revenue shares lie in the core and thus satisfy certain individual rationality conditions, (ii) the nucleolus of these games is polynomial-time computable, and (iii) one can implement budget-balanced group-strategyproof mechanisms, extracting people’s willingness to participate in the process. Finally, via simulations on real-world datasets, we conclude that with a fixed amount of revenue share, multi-level tree-based propagation will result in larger expected reach. Overall, we conclude that the tree-based multi-level propagation model for revenue sharing is more effective and more efficient than single-level or graph-based multi-level propagation models.

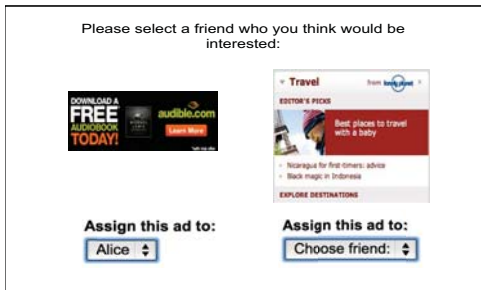


Figure 1: Users select most relevant friends.

2. RELATED WORK

Viral marketing over social networks has been studied for the purpose of influence maximization [12, 13], or revenue maximization [14, 15]. In the influence maximization models [12, 13], a person’s decision to buy a product is influenced by the set of other people who own the product. In the revenue maximization model [14, 15], people don’t simply adopt products, but rather must pay money to buy them. A person’s decision to buy a product is influenced by the set of

other people that own the product as well as the price at which the item is offered. On the subject of propagation of information and influence on social networks, various work has been done. In a recent paper [16], the authors argue that viral marketing would be more effective if a large number of ordinary users are picked as influencers. None of the above work however studies the effect of revenue sharing in incentivizing users of the social networks in maximizing the reach and effectiveness of online ads.

Recently, a social ad model considering user influence, called AdHeat, has been explored [4]. In this model, the advertising platform may diffuse hint words of influential users to others and then matches ads for each user with aggregated hints. They perform experiments on a real-world data set, and show that AdHeat outperforms the traditional relevance models by a large factor. Although this study shows the effectiveness of using social network information in online advertising, however they do not consider active propagation of ads by users of the social network.

The applications of the Shapley value to network and in particular Internet economics is also been of interest recently: from applications to peer-assisted services [17] to the settlement issue between Content, Eyeball, and Transit ISPs [18, 19, 20]. In a less recent work [21], authors propose that certain private information should only be disclosed by users if they get compensated fairly. The paper determines the value of private information in the context of online surveys, collaborative filtering and in general recommender systems by the Shapley value.

A related but different problem is the cost sharing of Steiner tree or multicast network design problem in cooperative game theory [22, 23]. It is not hard to see that cost function of a multicast tree for a subset of nodes is a submodular function, and this implies the existence of budget-balance cross-monotone cost sharing methods for this problem. This might seem related to our proof of supermodularity of the revenue function in the tree-based propagation model. However, in the multicast cost sharing context, the submodularity of the cost function holds for any general cost function on the edges while, it is not hard to show that supermodularity of the revenue function does not hold in the general case and we show that it holds in the case that the revenue share of each node in the graph is the same (Section 5).

3. MODEL

Let U be the set of users of a social network platform (SNP) \mathcal{P} . The SNP \mathcal{P} implements a viral marketing program helping an ad campaign or an online retailer, by giving incentives to its users to participate in marketing for the advertiser. To be more specific, the SNP suggests coupons, products or ads to a subset of its users, and asks them to refer a *limited* number of their friends for each. If their friends buy the coupon or product or click on the ad (or simply do anything through which the SNP gains revenue), the revenue gained would be shared with the referrers. If the friend also refers some of her friends a cascade of referrals would be initiated. In this paper, if a user adopts a product, buys a coupon or clicks on an ad or takes any action through which revenue is earned by the SNP (directly or through a cascade), we say that the node has become *active*. The nodes that had a role

in activating a node are called the *activators* of that node.

Looking at this model as a cooperative game, the set of players are denoted by U , where $N = |U|$. We call any nonempty subset $S \subseteq U$ a coalition of players. For each coalition S , we denote by $f(S)$ the worth function ($f : 2^{U \cup \{\mathcal{P}\}} \rightarrow R$), which measures the total revenue from an advertiser produced by the system when all players of this coalition S are active. Clearly the revenue from a subset S of users without \mathcal{P} is zero as any coalition needs the SNP to implement the marketing/advertising strategy. Let $f_u(S)$ denote the revenue of player u in the coalition S , we then have: $f(S) = \sum_{u \in S} f_u(S)$.

We suggest two models for revenue sharing if a node becomes *active*. In the first model, which is very similar to what happens commonly in referral-based marketing, the SNP shares the added revenue only with direct *activators* of that node. The other model suggests that the revenue be shared with the whole set of users who contributed in activating a node.

Before we get into specifics of our proposed models we define the Shapley value [5]:

DEFINITION 1 (SHAPLEY VALUE). *Shapley value of player u in coalition S is denoted as $\phi_u(S, f)$ and is computed as: $\forall u \in U, \phi_u(S, f) = \frac{1}{|S|!} \sum_{\pi \in \Pi} \Delta_u(f, S(\pi, u))$*

where Π is the set of all $|S|!$ orderings of S and $S(\pi, u)$ is the set of players preceding u in the ordering π . The Shapley value of player u can thus be interpreted as the expected marginal contribution $\Delta_u(f, S')$, where S' is the set of players in S preceding u in a uniformly distributed random ordering of S .

Shapley value of each player u in coalition S satisfies the three following axioms:

Axiom 1: Efficiency $\sum_{u \in S} \phi_u(S, f) = f(S)$ This axiom means that the total revenue share of each user should be equal to the total revenue gained in the coalition.

Axiom 2: Symmetry If for all $S' \subseteq S \setminus \{u, v\}$, $f(S' \cup \{u\}) = f(S' \cup \{v\})$ then $\phi_u(S, f) = \phi_v(S, f)$ This axiom states that players contributing equal amounts to a coalition should receive same amount of the revenue.

Axiom 3: Fairness For any $u, v \in S$, v 's contribution to u equals u 's contribution to v , or in other words $\phi_u(S, f) - \phi_u(S \setminus \{v\}, f) = \phi_v(S, f) - \phi_v(S \setminus \{u\}, f)$

We formulate the network of referrals by constructing a directed graph representing users influencing their friends in becoming *active*. More specifically, nodes in this graph are users of the SNP and there is an edge from user w to user u if w activates u . We also add a root node for the SNP and connect this root to all original users who become active. Given this graph, users who are on paths from the root to each node u share the revenue from the activity of this node. In fact, since there might be several users contributing in activation of a specific node, there are different ways to construct this graph. We will describe our models below.

Note that similar to other advertising and revenue sharing systems, such revenue sharing schemes should also be accompanied with a reasonable fraud detection and reputation system that can control for malicious user behavior. Also, It is important to note that in this model, a user is only allowed to refer an ad/product to a *limited* number of her

friends, therefore, increasing her incentive to pick the most relevant ones.

3.1 Single-level Propagation Model.

In the single-level propagation model the amount paid by the advertiser for a node becoming active is shared among the SNP and the set of its direct activators. Considering this model as a cooperative game, the players of this game are the SNP and the users who influence other nodes to become active. If k users refer a product/ad to user u and she becomes active resulting in revenue gain, the coalition would consist of these k users and the social network platform. The worth function is described in the following.

To be more precise, assume that the advertiser is willing to pay Q_u for each action, and the probability of an action by user u is p_u , i.e., the expected revenue of referring a product to u is $p_u Q_u$. In this setting, the revenue $f_u(S)$ from a user u given a subset $S \subseteq U \cup \{\mathcal{P}\}$ of users is computed as follows: if $\mathcal{P} \in S$ and there exists a user $u' \in S$ who is connected to user u and activates u , then $f_u(S) = p_u Q_u$, otherwise $f_u(S) = 0$. We should note that it is possible that k of u 's friends play role in activating u , therefore fair ways to share the revenue among the contributing users should be explored. We will discuss such fair methods in Section 4.

3.2 Multi-level Propagation Model.

The multi-level model generalizes the single-level model by sharing the revenue with all the users collaborating in a cascade of referrals. In other words, the multi-level model keeps track of the path of users activating their friends in the network and shares revenue with all of them. This way, users get credit from their friends' friends activation. This propagation can grow for multiple levels, which gives more incentive to users to make referrals because they will earn money not only from direct activations but also from all the chains they are part of. Note that in this model, the assumption is that users do not have an incentive to refer every item to all their friends, since that will result in losing their credibility and ruining their reputation.

Modeling this as a cooperative game, the players are the SNP and all the users who participate. The worth function would be as follows. In such a model, referrals are propagated through paths, and the expected revenue for a referral from a user u is $p_u Q_u$ if there is a path of users propagating the referral iteratively to user u , i.e., $f_u(S) = p_u Q_u$ if $\mathcal{P} \in S$ and there is a path of users $v_1, v_2, \dots, v_l \in S$ where v_1 originally has become active, and then each user v_i activates user v_{i+1} after receiving it from v_{i-1} , and finally v_l activates u , otherwise $f_u(S) = 0$.

There are various ways for the SNP to keep track of activation paths. These various ways would, in turn, impose different revenue sharing schemes. For example, the SNP may only keep track of the first user who makes a referral to each user. Alternatively, the SNP may keep track of all users who made referrals to another user. Different methods for keeping track of users making referrals can be divided into two main categories. Consider a set w_1, w_2, \dots, w_k activating user u .

Graph-based Model: The SNP may keep track of all

users who activated u , i.e., we may put edges from all nodes w_1, w_2, \dots, w_k to node u . **Tree-based Model:** We may put an edge only from one node w_i to user u , e.g. we may put an edge only from w_1 who is the first user who activated u encourage users to make referrals as early as possible.

For each of the above multi-level models, we can define a k -level propagation model in which the revenue sharing for a node u happens among at most k users on the path to u , i.e., the revenue from node u is shared only among the last k nodes on a path to node u . For example, in the k -level tree-based model, the revenue share for a user u is shared among k top parents of node u in the corresponding propagation tree. Throughout the paper, we mainly study the general multi-level model both for the graph-based and tree-based models, but all of our results hold for the k -level variant of these propagation models. In parts of the paper that we need to distinguish between different k -level models, we specify the k -level propagation model.

4. COMPUTING THE SHAPLEY VALUE

The challenge in the above mentioned models is to design a fair mechanism to share the added revenue with users that participated in the process. Shapley value as described in details in Section 3 not only ensures fairness, but also has other desirable properties. In what follows, we discuss computation of the Shapley value for the proposed models. We either show that Shapley value can be computed polynomially, or prove a hardness result and provide approximate solutions.

4.1 Single-level propagation model.

Assume that k users have activated user u . Here, we discuss simple fair ways to share the revenue gained among the contributing users and the SNP. One way is to just consider the first user who starts the propagation, in which case the revenue should only be divided between that first user and the SNP. Alternatively, we may share the revenue with all these k users¹. Consider user u , let K_u be the set of k_u users who activated user u , and let the revenue of user u becoming active be $p_u Q_u$. Then, the Shapley value revenue share of SNP for each user u is $\frac{k_u}{k_u+1} p_u Q_u$ and the Shapley value revenue share of each of these k_u users is $\frac{1}{k_u(k_u+1)} p_u Q_u$. Letting k_u be the set of users who have activated user u , and summing up the Shapley value revenue share for all users, the revenue share of the SNP in the single-level model is $\sum_{u \in U} \frac{k_u}{k_u+1} p_u Q_u$. Also letting A_i be the set of users who have been referred by user i , Shapley value for user i would be $\sum_{u \in A_i} \frac{1}{k_u(k_u+1)} p_u Q_u$.

4.2 Graph-based Multi-level propagation model

In this section, we show that computing the Shapley value in the graph-based multi-level model is computationally hard, and in fact is $\#P$ -hard. The proof is by reduction from a node variant of the NETWORKRELIABILITY problem, called NODEREIABILITY. Both problems are defined below:

NETWORKRELIABILITYProblem

Instance: Graph $G = (V, E)$, a rational failure probability $p(e)$ for each $e \in E$, nodes s and t .

Question: If edge failures are independent from each other, what is the probability that there exists a path from s to t in this graph?

NODEREIABILITYProblem

Instance: Graph $G = (V, E)$, a rational failure probability $p(v)$ for each $v \in V$, nodes s and t .

Question: If node failures are independent from each other, what is the probability that there exists a path from s to t in this graph?

The NETWORKRELIABILITY is known to be $\#P$ -complete, even for a fixed probability $p(e) = \frac{1}{2}$ for each edge e [24, 25]. Using this, it is not hard to show that NODEREIABILITY is also $\#P$ -hard by giving a reduction from the edge variant: Given an instance of the NETWORKRELIABILITY problem, construct an instance (s, t, G') of the NODEREIABILITY problem as follows: Let $V(G') = E(G) \cup \{s, t\}$, i.e., for each edge e in the graph G , put a node v_e in graph G' with $p(v_e) = p(e)$, and also put two nodes corresponding to s and t . Two nodes in G' are adjacent if their corresponding edges or nodes in G are adjacent. One can easily verify that the probability of having a path in a random sample of G in the NETWORKRELIABILITY problem is the same as the probability of having a path from s to t in the random sample of G' in the new instance of the NODEREIABILITY problem.

THEOREM 1. *Computing Shapley value in the multi-level graph-based propagation model is $\#P$ -complete.*

PROOF. Consider an instance of NODEREIABILITY problem as follows: Given a graph $G = (V, E)$ and two nodes s and t , and probability $p(v) = \frac{1}{2}$ on each node, compute the probability of having a path from s to t in a random graph constructed by including each node of G with probability $\frac{1}{2}$. From this instance, we construct an instance of the Shapley value computation in the multi-level graph-based model. The propagation graph G' is the same as graph G with an additional node v and two edges (s, v) and (v, t) , i.e., $V(G') = V(G) \cup \{v\}$ and $E(G') = E(G) \cup \{(s, v), (v, t)\}$. Now consider the revenue share of nodes in G' toward node t , that is the revenue shares toward $p_t Q_t$. We claim that the total revenue share of nodes other than v in G' is $p_t Q_t$ times the probability of having a path from s to t in a random graph where each node of G is present with probability $\frac{1}{2}$. To see this, note that in computing the Shapley value revenue shares of nodes in $V(G') \setminus \{v\}$, the probability that each node $u \in V(G)$ appears before v in a permutation is $\frac{1}{2}$. Thus the probability that each node is before v is $\frac{1}{2}$ and is independent of any other node appearing before v . Also, if a path from s to t appears completely before v in a permutation, the marginal value of node t which is $p_t Q_t$ goes to one of the nodes in $V(G') - \{v\}$. Therefore, the total revenue share of nodes in $V(G') - \{v\}$ in the multi-level graph-based propagation model over G' is equal to the probability of having a path from s to t in a random subgraph of G where each node is present with probability $\frac{1}{2}$. Thus if we

¹Considering the influence model known as the threshold model, it is reasonable to assume that all referrers should receive some credit.

can compute the Shapley value revenue shares, we can solve the NETWORKRELIABILITY problem which is #P-hard. \square

4.2.1 Approximating Shapley Value

In light of the above hardness result, we design an algorithm based on sampling to approximate Shapley value for the graph-based multi-level model. It can be observed that by simply using polynomial number of samples we can compute Shapley values approximately in this general model. For completeness, we sketch the algorithm and the proof in the Appendix 7.

THEOREM 2. *If $\phi_u(S, f_v) > \frac{P_v Q_v}{n^3}$, then we can compute $\phi_u(S, f_v)$ within factor $(1 \pm \epsilon)$ with high probability (i.e., probability $1 - o(1)$), in time polynomial in $\frac{1}{\epsilon}$ and n . Otherwise, if $\phi_u(S, f_v) \leq \frac{P_v Q_v}{n^3}$, one can approximate it within multiplicative factor and $\frac{P_v Q_v}{n^3}$ additive error, in time polynomial in $\frac{1}{\epsilon}$ and n .*

4.3 Tree-based propagation model.

Here, we show that the Shapley value revenue share of each user in the tree-based model can be computed easily.

LEMMA 3. *In the multi-level tree-based propagation model, the Shapley value revenue share of each user can be computed in time $O(n^2)$.*

5. SUPERMODULARITY

In this section, we observe a main advantage of the tree-based propagation model compared to the graph-based propagation model. In particular, we show that the revenue function for the multi-level tree-based model described above is supermodular, and this implies various nice properties of the Shapley value revenue shares for the tree-based propagation model. For example, this shows individual rationality of these revenue shares for the corresponding cooperative game. We first prove the supermodularity and then switch to summarizing the corollaries.

As we have explained at the end of Section 2, it is important to note that although this result might seem related to the multi-cast cost sharing problem, the results are different. Before stating the proof of supermodularity for the tree-based model, we observe that this property does not hold for the graph-based propagation model, and even for the single-level propagation model with uniform valuations for the revenue shares. To see this, consider the following example:

EXAMPLE 1. *Consider a single-level model with 4 users A, B, C, D in which all 3 users A, B, C make referrals to user D and D becomes active. Let the revenue of each user be 1. In this case, the value of each subset of size 2 including s and one of A, B or C is 2 (since using A, B , or C , the path from D is formed). Also $f(s) = 0$, $f(s, A, B) = 3$, since there are three nodes reachable from s each with revenue 1. This example violates supermodularity as follows: $f(s) = 0$, $f(s, A) = 2$, $f(s, B) = 2$ and $f(s, A, B) = 3$ and $f(s, A, B) - f(s, B) = 3 - 2 = 1 < 2 = 2 - 0 = f(s, A) - f(s)$.*

THEOREM 4. *The revenue function in the tree-based multi-level propagation model with uniform valuation for all users $p_u Q_u = P$ is a supermodular set function.*

PROOF. A potential way to show that function f is supermodular is by showing that the set function f_u for any user u is supermodular. However, it is not the true in this case, i.e., for some users u , the set function f_u might not be supermodular. Nevertheless, we can show that the summation $f(S) = \sum_{u \in U} f_u(S)$ is supermodular. In the tree-based propagation model, for a subset $S \subset U \cup \{s\}$, $f(S) = 0$ if $s \notin S$, and otherwise, $f(S)$ is equal to $|T(S)|P$ where $T(S)$ is the maximal connected subtree rooted at s with all internal nodes in S . In other words, $f(S)$ is proportional to the number of nodes that are connected to s using a path whose all internal nodes are in S . Knowing $f(S) = P|T(S)|$ in the uniform valuation model, it is sufficient to prove that $|T(S)|$ is supermodular. We do so by verifying the supermodularity property of f by proving that for any two subsets $A \subset B$ and any element $i \notin B$,

$$|T(B \cup \{i\})| - |T(B)| \geq |T(A \cup \{i\})| - |T(A)|.$$

Letting $\Delta_i(B) = T(B \cup \{i\}) - T(B)$ and $\Delta_i(A) = T(A \cup \{i\}) - T(A)$, it is sufficient to prove that $\Delta_i(A) \subseteq \Delta_i(B)$. To show this, we consider two cases:

Case 1: i 's parent is not in $T(A) \cap A$. In this case, adding i to A does not change $T(A)$, and thus $\Delta_i(A) = \emptyset \subseteq \Delta_i(B)$.

Case 2: i 's parent is in $T(A) \cap A$. In this case $\Delta_i(A)$ contains all non-root nodes of the maximal connected subtree rooted at i with all internal nodes in A . In this case, i 's parent is also in $T(B) \cup B$, and $\Delta_i(B)$ contains all non-root nodes of the maximal connected subtree rooted at i with all internal nodes in B . Now since $A \subseteq B$, the corresponding connected maximal subtree in the induced graph of B is larger than the maximal subtree rooted at i in A . The result follows from the above case analysis, as it shows that $\Delta_i(A) \subseteq \Delta_i(B)$, and thus $|\Delta_i(A)| \leq |\Delta_i(B)|$ which implies supermodularity of f . \square

The supermodularity of revenue shares, in turn, implies individual rationality, computability, and incentive compatibility results for the revenue shares based on the tree-based propagation model with uniform valuations. In the appendix 7 we list three corollaries of this property.

6. SINGLE-LEVEL VS. MULTI-LEVEL

In the previous sections, we argued for an advantage of using tree-based propagation model compared to a general graph-based propagation model. In this section, we compare single-level and multi-level propagation models, and show which revenue sharing strategy is more effective in maximizing the spread of viral marketing or advertising campaign.

Let M be the total revenue share for each new user who gets the referral, i.e., $M = pR$ where p is the probability that a new user becomes active, and R is the amount that the online retailer or the advertiser pays as the total revenue share with each user. We observe that there exists a direct relation between the amount of revenue shared with users and the amount of spread of the advertising campaign. Intuitively, the more revenue share there is, the more probable the spread is. Here, we observe that other than the parameter M , the revenue sharing scheme also plays an important role in the expected reach of the viral marketing method. Intuitively, for multi-level revenue sharing schemes, the potential gain of each user for making referrals is more than the potential gain for single-level sharing schemes. The rea-

son is that the user not only gains a revenue share from her immediate neighbors, but also from her neighbors of neighbors and so on. This in turn increases the probability of making referrals by users, and thus it may result in higher expected reach of the marketing for multi-level revenue sharing schemes.

We model the potential reach of a revenue sharing strategy by simulating a random propagation process on real-world networks, and reporting the simulation results. In order to simulate this process, we consider the following model over a network: Each each user u has a random threshold t_u , where t_u is chosen uniformly at random from $[0, 1]$. Before making a decision to make a referral or not, each user computes her potential gain, $Potential(u)$, and makes referrals if $Potential(u) \geq t_u$. This propagation model is inspired by the probabilistic threshold model that is widely studied as a model for viral marketing [13]. An important feature of our model is the way we compute the potential gain of users. In fact, the main difference in various revenue sharing policies is the way users compute their potential gain. If we use a k -level tree-based propagation model for a large k , there is a larger potential revenue from referrals (at the beginning of the propagation process). In such a setting along with a k -level tree-based revenue sharing, a user u may get a revenue share $\frac{M}{k}$ (or $\frac{M}{t}$ for some $t \leq k$) for each new user who gets a referral from u , and also u may get a revenue share of $\frac{M}{k}$ from each new user who becomes active through u , and so on. As a result, the total potential gain of user u for making referrals is proportional to the potential number of new users within a distance k of user u who became active through u . At the beginning of the propagation process, this potential is larger for each node u , but as time goes on, more people hear about it through other friends, and the chance that a new user is informed by user u becomes lower. The more k is, the more potential gain users have at the beginning, and thus there is a higher chance that they start propagating referrals. On the other hand, for a larger k , after a fixed number of steps, more people have already heard about the product/ad, and the potential gain of users to propagate it to new users is lower. Because of this tradeoff, using different propagation models will result in different expected number of users who get the referral.

We simulate the above process for different k -level tree-based propagation models for several networks, and report the total expected number of users who end up getting the referral using different propagation models. We denote this expected number of users who hear about a product using a k -level tree-based model by $E(k)$. As we will explain in details, in all simulations, we observe that this expected number $E(k)$ increases as k increases.

Data. To evaluate the performance of our models we tested them on five large real networks. Due to space constraints, we report one of them in this section of the paper as the behavior for all the networks were similar. The network presented here is a who-trusts-whom network of Epinions.com which is a consumer review website. Users of the site can explicitly indicate whether they trust another member or not. The graph contains a node for each user and there is an edge from each user to the other users that she trusts. The graph has 75879 nodes and 508,837 edges between the nodes [28].

Table 1: Epinions dataset properties.

Nodes	Edges	Avg. Degree	Diameter	largest CC
75879	508837	13.41	13	32223

Results. The performance of the k -level propagation model is depicted in Figure 2 for the Epinions network. The simulation is done for $k = 1, \dots, 20$, for $M = 0.3$ and $M = 0.5$. The plots indicate the number of users that hear about the product for each k -level simulation. As the plots show the number increases as k increases up to 13 levels for the Epinions network. At level $k = 13$, the network is saturated.

We observe that the saturation level is close to the diameter of the graph. This implies that implementing a k -level model beyond the diameter will not improve the number of informed users. The 90 percent effective diameter of the graph is 5.8. We also observe that the relative increase in the number of users getting the referral is at maximum between levels 4, 5 and 6. We also find that, especially in the case where the total revenue share M is smaller, having a k -level propagation model is even more effective to trigger users initiate propagation. This observation can be interpreted intuitively as follows: when M is small, the single-level model might not be incentivizing enough for the users to opt in, however, by increasing the levels the potential increases which gives more incentive to users to take part in this process.

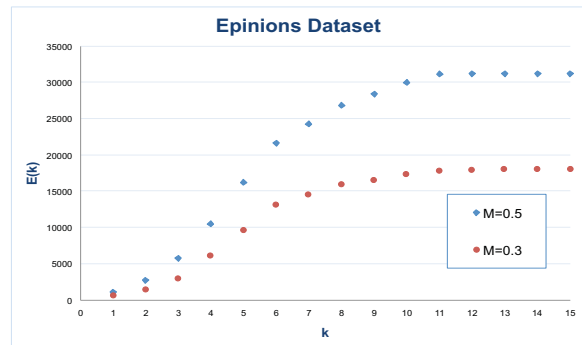


Figure 2: Epinions.com

7. CONCLUSION

In this paper, we have developed multi-level revenue sharing schemes for viral marketing over social networks. For the proposed models, we develop results on the computational complexity, individual rationality, and potential reach of employing the Shapley value as a revenue sharing scheme. Our results indicate that under the multi-level tree-based propagation model, the Shapley value is a promising scheme for revenue sharing, whereas under other models there are computational or incentive compatibility issues that remain open. We also assess the effectiveness and potential reach of the single-level and k -level tree-based models through simulations, and our findings show that using a k -level tree-based model has higher potential for increasing the spread over the social network.

8. REFERENCES

- [1] <https://networkchallenge.darpa.mil/Default.aspx>.
- [2] DARPA Network Challenge Project Report. February 16, 2010. <http://www.eecs.harvard.edu/cs286r/papers/ProjectReport.pdf>
- [3] G. Pickard, I. Rahwan, W. Pan, M. Cebrian, R. Crane, A. Madan, A. Pentland. Time Critical Social Mobilization: The DARPA Network Challenge Winning Strategy. arXiv:1008.3172v1
- [4] H. Bao, E. Y. Chang AdHeat: an influence-based diffusion model for propagating hints to match ads. WWW 2010: 71-80.
- [5] L. Shapley. A Value for n-Person Games Contributions to the Theory of Games II, volume 28 of Annals of Mathematics Studies, pages 307-317. Princeton University Press, 1953.
- [6] L. Shapley "Cores of Convex Games, International Journal of Game Theory 1, 11-26, 1971.
- [7] J. Kuipers. A polynomial time algorithm for computing the nucleolus of convex games. Technical Report M 96-12, Maastricht University, 1996.
- [8] J. Kuipers, U. Faigle, W. Kern,. On the computation of the nucleolus of a cooperative game, International Journal of Game Theory, Springer, vol. 30(1), pages 79-98, 2001.
- [9] J.J. Brown, P. H. Reingen, Social ties and word-of-mouth referral behavior. J. Consumer Research 14, 3, 350-362, 1987.
- [10] A. DeBruyn, G. Lilien A multi-stage model of word-of-mouth through viral marketing. International Journal of Research in Marketing, 25(3), 143-225. 2008.
- [11] J. Leskovec, L. A. Adamic and B. A. Huberman. The dynamics of viral marketing, TWEB 1, 1, 2007.
- [12] P. Domingos, M. Richardson. Mining the network value of customers. In Proc. KDD 2001. 57-66.
- [13] D. Kempe, J. Kleinberg, E. Tardos, Maximizing the spread of influence through a social network. In Proc. KDD 2003, pp. 137-146. ACM, New York (2003)
- [14] J. Hartline, V. S. Mirrokni, M. Sundararajan. Optimal marketing strategies over social networks, WWW, 2008, 189-198.
- [15] H. Akhlaghpour, M. Ghodsi, N. Haghpanah, H. Mahini, V. S. Mirrokni, and P. Nikzad. Iterative pricing over social networks. In Proc. WINE 2010.
- [16] E. Bakshy, J. M. Hofman, W. A. Mason, D. J. Watts Everyone's an influencer: quantifying influence on twitter. WSDM 2011: 65-74.
- [17] V. Misra, S. Ioannidis, A. Chaintreau, and L. Massoulié. Incentivizing Peer-Assisted Services: A Fluid Shapley Value Approach. In Proc. ACM Sigmetrics, 2010.
- [18] R. T. B. Ma, D. Chiu, J. C. Lui, V. Misra, and D. Rubenstein. Internet Economics: The use of Shapley value for ISP settlement. In Proc. ACM CoNEXT, 2007.
- [19] R. T. Ma, D. Chiu, J. C. Lui, V. Misra, and D. Rubenstein. On cooperative settlement between content, transit and eyeball internet service providers. In Proc. of ACM CoNEXT, 2008.
- [20] Y. Cheung, D. M. Chiu, and J. Huang, Can Bilateral ISP Peering Lead to Network-wide Cooperative Settlement? IEEE International Conference on Communications and Networks, 2008
- [21] J. Kleinberg, C. Papadimitriou, P. Raghavan. On the Value of Private Information. Proc. 8th Conf. on Theoretical Aspects of Rationality and Knowledge, 2001.
- [22] K. Jain and V. V. Vazirani. Applications of approximation algorithms to cooperative games. STOC, 2001.
- [23] J. Feigenbaum, A. Krishnamurthy, R. Sami, and S. Shenker. Hardness Results for Multicast Cost Sharing Theoretical Computer Science 304 (2003).
- [24] C. J. Colbourn. The Combinatorics of Network Reliability. Oxford University Press, Inc., New York, NY, USA, 1987.
- [25] R. Karp and M. Luby. Monte carlo algorithms for the planar multi-terminal network reliability problem. Journal of Complexity, 1:45-64, 1985.
- [26] M. Newman. The structure of scientific collaboration networks. Proc. Natl. Acad. Sci. 98, 2001.
- [27] J. Leskovec, J. Kleinberg and C. Faloutsos. Graph Evolution: Densification and Shrinking Diameters. ACM TKDD, 1(1), 2007.
- [28] M. Richardson and R. Agrawal and P. Domingos. Trust Management for the Semantic Web. ISWC, 2003.
- [29] H. Moulin and S. Shenker, 1992. "Serial Cost Sharing," Econometrica, Econometric Society, vol. 60(5), pages 1009-37.

Appendix

Proof of Theorem 2

PROOF. Consider the following algorithm based on sampling: First, generate $m = O(\frac{n^7}{4\epsilon^2})$ random permutations of the players 1 to n . For each player u , compute the the marginal contribution of player u in each of these m permutations. Let the revenue share of each player u be equal to the average of its marginal contributions over these m permutations. Let X_i be the marginal contribution of player u in trial i of the sampling algorithm. Let $X^v(u)$ be the random variable which is the output of the sampling algorithm using $m = O(\frac{n^7}{4\epsilon^2})$ samples, i.e., $X^v(u) = \frac{\sum_{i=1}^m X_i}{m}$. The expected value of $X^v(u)$ is equal to $\phi_u(S, f_v)$ by definition, i.e., $E[X^v(u)] = \phi_u(S, f_v)$. Let $\sigma^2_{X^v(u)}$ be the standard deviation of the sample mean, and σ_i be the standard deviation of X_i .

Using Chebyshev's inequality, we have:

$$\Pr[|X_u^v - \phi_u(S, f_v)| \geq \epsilon \phi_u(S, f_v)] \leq \frac{\sigma^2_{X^v(u)}}{\epsilon^2 \phi_u(S, f_v)^2},$$

It is not hard to see that

$$\sigma^2 \leq \frac{(X_{i_{max}} - X_{i_{min}})^2}{4},$$

where $X_{i_{max}}$ and $X_{i_{min}}$ are the minimum and maximum marginal value for u in any permutation of elements of S . Since marginal value is non-negative and at most $P_v Q_v$, thus $X_{i_{max}} \leq P_v Q_v$ and $X_{i_{min}} \geq 0$, and therefore, $\sigma_i^2 \leq \frac{(P_v)^2 (Q_v)^2}{4}$, or $\sigma_i \leq \frac{P_v Q_v}{2}$, and thus

$$\sigma^2_{X^v(u)} = \frac{\sigma_i^2}{m} \leq \frac{P_v^2 Q_v^2}{4m} = \frac{\epsilon^2 P_v^2 Q_v^2}{n^7}.$$

Using the above fact, if $\phi_u(S, f_v) \geq \frac{P_v Q_v}{n^3}$, we can rewrite the Chebyshev's inequality as follows:

$$\begin{aligned} \Pr[|X_u^v - \phi_u(S, f_v)| \geq \epsilon \phi_u(S, f_v)] &\leq \\ \frac{\sigma^2_{X^v(u)}}{\epsilon^2 \phi_u(S, f_v)^2} &\leq \\ \frac{n^6 \epsilon^2 P_v^2 Q_v^2}{n^7 \epsilon^2 P_v^2 Q_v^2} &= \frac{1}{n}, \end{aligned}$$

as desired. For the case of $\phi_u(S, f_v) \leq \frac{P_v Q_v}{n^3}$, one can use the same Chebyshev's inequality with an additional $\frac{P_v Q_v}{n^3}$ additive error.

□

The above method can be easily generalized for computing revenue shares in the k -level graph-based propagation model approximately. The only difference in the proof and in the algorithm is in computing the marginal contribution of adding each player in the permutation.

Proof of Lemma 3

PROOF. The revenue function f can be written as the sum of the revenue functions from each user u , i.e., $f(S) = \sum_{u \in U} f_u(S)$, and $f_u(S)$ for the tree-based propagation model is defined in Section 3. Thus in order to compute the Shapley value revenue share of each node in a coalition S , we can compute the revenue share for each user u . In order to compute the revenue share of each user from user u , we note the set of users who gets some share of the expected revenue $p_u Q_u$ generated from u are the nodes on the path from the root s to node u in the tree. Therefore, we focus on the path $s = u_0, u_1, u_2, \dots, u_n = u$ of n users from the root s to the node u in the tree, and compute the Shapley value of user u_k on this path. There is also a root node s for SNP which is connected to u_1 . It is not hard to see that in any random permutation of $\{u_0, \dots, u_{n-1}\}$, the marginal revenue of adding an element to the list of elements is $p_u Q_u$ if and only if this elements appears at the end of the permutation which happens with probability $\frac{1}{n}$. Therefore, the revenue share of each user $u \in u_0, u_1, \dots, u_{n-1}$ from user u is $\frac{p_u Q_u}{n}$. Hence, the total Shapley value of a node v with a subtree T_v under v is equal to $\sum_{u \in T_v} \frac{p_u Q_u}{D_u}$, where D_u is the depth of user u in the tree. □

Corollaries of Theorem 4

First of all, this supermodularity implies a desired individual rationality property of the Shapley value revenue shares, since the Shapley value revenue shares lie in the core [6].

COROLLARY 5. *For the tree-based propagation model with uniform valuations, the core of the corresponding cooperative game is non-empty and can be computed in polynomial time. In particular, the Shapley value revenue shares for the tree-based propagation model with uniform valuations lie in the core of the cooperative game.*

Moreover, the above supermodularity result implies existence of a non-empty nucleolus for the corresponding game [7, 8].

COROLLARY 6. *For the tree-based propagation model with uniform valuations, the nucleolus of the corresponding cooperative game is non-empty and can be computed in polynomial time.*

Moulin Mechanisms. One way to deploy a revenue sharing scheme for ad propagation or referral marketing is to run a mechanism asking users' willingness to opt in the propagation scheme. In order to run such a mechanism, the SNP can have users bid the minimum value they need to receive in order to opt in the referral marketing system. In such settings, given the information the SNP has about the network, it can estimate a potential revenue function $f(S)$ for each subset S of users. Now the SNP encounters the following mechanism design problem: Given a set of bids b_i from each user i , what mechanism should we use to decide on a set S of users to help the provider and the amount to pay them? A desired goal is to design a mechanism that satisfies the following properties: (i) Individual rationality, i.e., each user should be in set S only if the SNP can pay him/her more than his bid b_i . Also each user not in S should get 0, (ii) Truthful (or group-strategyproof): Each user (or any group of users) should have an incentive to reveal their true value b_i as their bid, and (iii) budget-balance, i.e., the total revenue $f(S)$ of the set should be divided in some way among users and the provider.

One way to design a mechanism with the above properties is via finding a *cross-monotonic* revenue sharing method, i.e., finding a function $v : U \times 2^U \rightarrow R$, where for each two subset $S \subseteq T \subseteq U$ of users and each user $i \notin T$, we have $v(i, S) \leq v(i, T)$. A revenue sharing method is *budget-balanced* for a revenue function f , if for any subset $S \subseteq U$, $\sum_{i \in S} v(i, S) = f(S)$. Given a budget-balanced cross-monotonic revenue sharing method for a revenue function f , one can design Moulin mechanisms satisfying all the above properties. It is also known that for supermodular revenue functions f , there exists a budget-balance cross-monotonic revenue sharing method v for f [29]. As a result, we get the following:

COROLLARY 7. *Given a revenue function f based on the tree-based model, one can design a mechanism satisfying individual rationality, group-strategyproofness, and budget-balance properties.*