How Do We Find Early Adopters Who Will Guide a Resource Constrained Network Towards a Desired Distribution of Behaviors?

KAUSHIK SARKAR, Arizona State University
HARI SUNDARAM, Arizona State University

We identify influential early adopters that achieve a target behavior distribution for a resource constrained social network with multiple costly behaviors. This problem is important for applications ranging from collective behavior change to corporate viral marketing campaigns. In this paper, we propose a model of diffusion of multiple behaviors when individual participants have resource constraints. Individuals adopt the set of behaviors that maximize their utility subject to available resources. We show that the problem of influence maximization for multiple behaviors is NP-complete. Thus we propose different heuristics based on node degree, Influence Weight and Immediate Adoption to select early adopters. We evaluate the effectiveness under three metrics: unique number of participants, total number of active behaviors and network resource utilization. We also propose heuristics to distribute the behaviors amongst the early adopters to achieve a target distribution in the population. We test our approach on synthetic and real-world topologies with excellent results. Our heuristics produce 15-51% increase in resource utilization over the naïve approach.

Categories and Subject Descriptors: J.4 [**Social and Behavioral Sciences**]; I.6.3 [**Simulation and Modeling**]: Applications

General Terms: Algorithms, Theory, Experimentation

Additional Key Words and Phrases: Behavior diffusion, seed selection, social diffusion, social network, viral marketing

ACM Reference Format:
Kaushik Sarkar, and Hari Sundaram, 2013. How Do We Find Early Adopters Who Will Guide a Resource Constrained Network Towards a Desired Distribution of Behaviors? *ACM Trans. Knowl. Discov. Data.* V, N, Article A (January YYYY), 27 pages.
DOI : http://dx.doi.org/10.1145/0000000.0000000

1. INTRODUCTION

This paper investigates how costly, multiple behaviors spread in social networks where individual have resource constraints. These constraints could include time, money or any material resource relevant to adopting the behavior. The problem is challenging because the problem of finding early adopters or seeds is known to be computationally intractable. We develop heuristics to find early adopters that maximize multiple behavior adoption over resource constrained social networks.

Many of the pressing challenges facing contemporary society concern sustainability and public health. For example, how can sustainable behaviors—such as reducing individual energy consumption—be encouraged? How can participation in activities that reduce overall healthcare costs—such as compliance with preventive care routines and leading healthy lifestyles—be supported? These questions are termed as *collective action* problems in the social sciences [Ostrom et al. 1999].

We are motivated by collective action problems to answer questions such as: how does a person's limited resources, including time, money or lack of tangible resources like a car or a bicycle, affect how she participates in real-world activities? A person interested in adopting a behavior (e.g. taking newspapers to a recycling station or voting) may fail to do so, due to lack of resources. A person may be interested in adopting multiple behaviors, but each behavior has a cost. Current models of

DOI : http://dx.doi.org/10.1145/0000000.0000000

behavior adoption lack the idea that individuals may have significant resource constraints that preclude them from successfully adopting behaviors in which they are interested. Resource constraints not only limit individual participation, but also shape how behaviors spread in a network. We are interested in maximizing the use of resources available in a social network towards adopting a set of behaviors.

In this paper, we develop a model of multiple behavior diffusion that captures the complex dynamics of multiple behavior adoption in resource constrained networks. Our work is the first of its kind, to the best of our knowledge to study the influence of individual resource constraints on multiple, costly behavior adoption. In our model, behaviors have associated costs and utilities that are independent of the individual. Mindful of the work by [Aral et al. 2009] and [Shalizi and Thomas 2011], individuals in our model evaluate a utility function for each behavior that combines intrinsic interest and social signals. An individual adopts a behavior when she receives a social signal of sufficient strength, the behavior is of high utility and when she has the resources to do so. We use three metrics: unique number of participants, number of behaviors in the network and expected resource utilization. Then, we identify two problems: which seeds to select, and what behaviors they need to adopt to maximize each of the three metrics.

We develop several heuristics to identify seeds that maximize the expected resource utilization in the network. These heuristics are necessary since we show that the problem of seed selection to maximize expected utilization is NP-complete. We also develop a heuristic to address a simple question: how do we distribute behaviors amongst the seeds to achieve a target behavior distribution? We test our approach on synthetic and real-world topologies with excellent results. We show that two heuristics that evaluate Influence Weight and Immediate Adoption provide very good solutions to the seed selection problems. We show that setting the seed behavior distribution to be proportional to the target behavior distribution produces excellent results. Our heuristics produce 15-51% increase in resource utilization over the naïve approach.

The rest of the paper is organized as follows. In the next section we review the relevant literature. In Section 3 we formally define our behavior diffusion model. In Section 4 we define the seed selection problem, present different heuristics and compare their performance. In Section 5 we discuss different behavior distribution strategies and present simulation results. In Section 6 discuss how to achieve a target distribution of behaviors in the network. In Section 7 we discuss open issues and extensions. Finally, we present our conclusions in Section 8.

2. RELATED WORK

The literature on social diffusion processes in vast. Diffusion of innovation was studied in [Rogers 1962]. Subsequent work on modeling such processes through epidemiological models was carried out by [Bass 1969] which became well known as the *Bass Diffusion Model* in the management sciences community. Another approach of explaining such processes through threshold based models were made popular by [Granovetter 1983]. In recent years the study of diffusion modeling has seen substantial development. [Watts 2002] presents a threshold based model of global cascades and analyzes why certain networks may appear to be "robust" yet turn out to be fragile against such cascades.

Works of [Domingos and Richardson 2002] and [Kempe et al. 2003] initiated the study of the computational problem of seed selection in the context of a "viral" social diffusion process. [Kempe et al. 2003] formalized the algorithmic problem for *Independent Cascade* and *Linear Threshold* models, proved the intractability results and provided a greedy approximation algorithm to the problem. However the approximation algorithm incurred a huge computational cost in practice. [Leskovec et al. 2007b] came up with CELF technique to reduce the simulation cost of the algorithm. [Chen et al. 2009] tried to address that problem by coming up with computationally cheap heuristics that match the performance of the approximation algorithm. A complementary data mining perspective of inferring the diffusion model parameters from the past interactions the was taken by [Saito et al. 2008], [Goyal et al. 2010] and [Mathioudakis et al. 2011].

Our work is informed by these literature, but is markedly different in a number of aspects. Much of the existing literature is concerned with diffusion of a single influence, while simultaneous diffusion of multiple influences is a more realistic scenario. Moreover none of these works take constraint on user resources into account and thereby does not apply directly to our problem of long term adoption of behaviors. [Bharathi et al. 2007] and [Carnes et al. 2007] discuss the problem of multiple competing influences, but they also do not incorporate the resource constraints or the utility maximizing behavior of social agents into their models. To the best of our knowledge the present work is the first investigation of the seed selection problem for multiple behavior diffusion in a resource constrained social network.

Next we mention a few interesting critiques of the study of social diffusion. The study of social diffusion and information contagion has met with its fair share of criticism. [Aral et al. 2009] argues that in their observational study more than 50% of the perceived behavior contagion can be attributed to homophily instead of social influence. However [Shalizi and Thomas 2011] have shown that homophily and social influence are generically confounded in social diffusion processes and it in general not easy to distinguish between the two effects. We have tried to take this observation into account while developing our model where. In our model the diffusion process is not exclusively driven by the social influence effects but an individual's intrinsic characteristics (i.e. resource constraint) also plays an important part in the adoption decision making. An important critique of the study of modeling mass adoption through epidemic like contagion models is put forward by [Goel et al. 2012]. They have analyzed a number of real world product diffusion events and found out that in most of the cases the diffusion stops within one degree of the initial adopting seed, thus drawing a sharp contrast with multi-step person to person contagion of influence. It is however unclear if their findings will generalize to collective action problems, where participating individuals express interest in adopting a behavior (e.g. eating healthy foods, take a flu shot). As we discuss in our open issues section (Section 7), there may be confounding effects between the communication problem and the decision making problem. That is, behavioral diffusion can halt in the presence of poor communication.

3. MODELS OF BEHAVIOR ADOPTION UNDER RESOURCE CONSTRAINT

In this section we introduce our behavior diffusion models. First we will describe our model of multiple behavior diffusion in a resource constrained network in the most general form. Next we will present a simplified version of the general model which will provide us with some analytical insight into the algorithmic problems to be defined in the next section. Then we will introduce metrics, including resource utilization, unique participation and number of behavior adoptions to evaluate the behavior adoption process.

3.1. A Model of Multiple Behavior Diffusion in a Resource Constrained Social Network

We now describe the model for each user, the properties of each behavior and the behavior adoption process. Conceptually our behavior adoption model can be described as follows - an individual adopts a new behaviors if the behavior has some value to him, or he has some interest in the behavior (intent), many of her friends have adopted the behavior (social signal), and she has enough available resource to pursue it (resource). Her payoff from adopting that behavior will be determined by her intent and social signal. Figure 1 and **??** shows the idea in a simple way. Next we try to describe this idea formally.

We represent the social network with an undirected graph $G = (V, E)$. Each node $v \in V$ of the graph G represents an individual and an edge $e \in E$ between two nodes indicate a social relationship between the two individuals.

We wish to spread k behaviors in the social network. Each behavior i has an associated cost c^i and a utility u^i. The cost refers to the cost of adoption and the utility refers to the intrinsic utility gained by an individual by adopting this behavior. In a simplification, we assume that both the cost c^i and the utility u^i of behavior i are intrinsic to the behavior and independent of the individual who adopts the behavior. Without loss of generality, we assume that $0 \leq c^i, u^i \leq 1$.

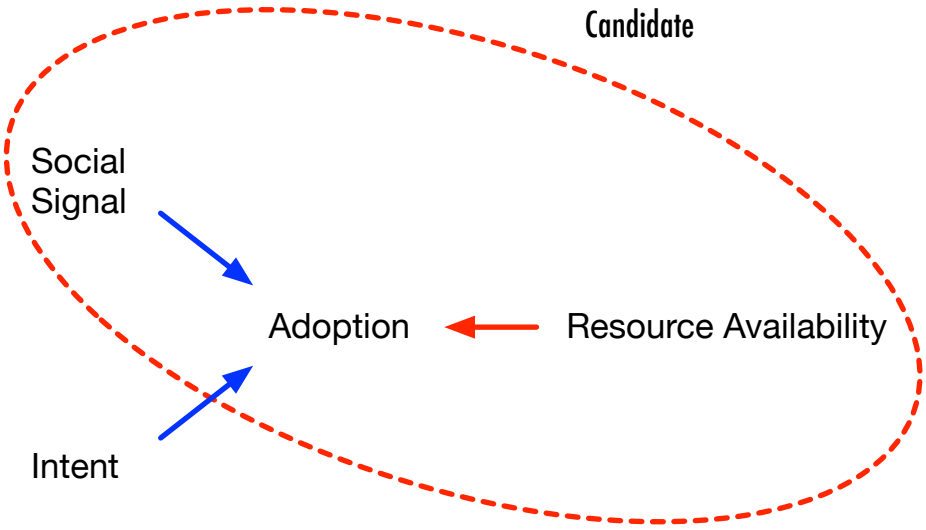


Fig. 1: *Sufficient social signal, intent and resource make an individual a candidate for adoption of a behavior.*

Individuals are resource constrained: an individual may have limited time, money or may not possess other material resources to adopt a behavior. Therefore, we assign a *fixed* resource $r(v)$ for each individual $v \in V$ towards adopting behaviors. The resource satisfies $0 \leq r(v) \leq 1$. For example, if we assume that individuals' resources are independent and identically distributed then the resource value $r(v)$ can be assumed to be obtained from a uniformly distributed random variable $U(0,1)$. Let $N(v)$ denote the set of neighbors of v in the network. Then we assume that a neighboring node u asserts a social influence on node v with weight $1/|N(v)|$.

An individual will adopt a behavior i when she receives a strong social signal, has the resources to do so and when the behavior is of a sufficiently high utility. An individual v adopts a behavior i when the social signal exceeds a threshold $\theta^i(v)$, where $0 \leq \theta^i(v) \leq 1$. We assume that each individual v has a different, fixed, threshold for each behavior, and that each threshold is obtained from a uniformly distributed random variable $U(0,1)$. An individual v can adopt a behavior i provided the cost c^i is less than $r(v)$, the available resources. The payoff $p^i(v)$ for a behavior i is defined as the weighted sum of the intrinsic utility u^i and the local network utility $l^i(v)$. That is, $p^i(v) = wu^i + (1-w)l^i(v)$. Where, w denotes the relative weight of the intrinsic utility and where $l^i(v)$ denotes the sum of influence weights—the social signal—exerted on v by its neighbors who have adopted behavior i. If there are multiple behaviors that can be adopted, an individual will adopt a subset that maximizes payoff.

Let us examine the diffusion of behavior over time, to illuminate the key ideas. The process takes place over discrete epochs[1]. We assume each node is aware of the behaviors adopted by her neighbors. The individual v first identifies all candidate behaviors. A behavior j is a candidate

[1]Notice that while actions in a network are asynchronous, we can choose an appropriate time granularity for analysis to assume synchronized decision making.

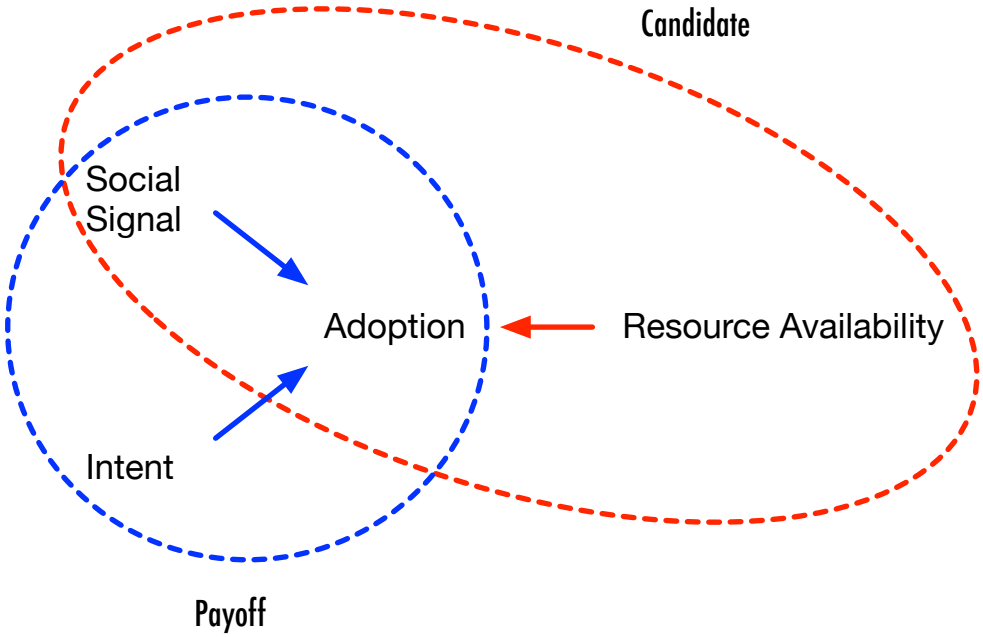


Fig. 2: *Payoff for adopting a behavior comes from an individual's intent and social signal.*

to be adopted if two conditions hold. First the social signal strength for behavior j must exceed the threshold for that behavior at node v—$l^j(v) \geq \theta^j(v)$. Second, the individual v must have the resources to adopt the behavior—$r(v) \geq c^j$. The first condition is the familiar Linear Threshold (LT) model [Kempe et al. 2003]. Since there are multiple behaviors, the individual v chooses a set of behaviors that will maximize the total payoff (i.e. the sum of payoffs $\sum_i p^i(v)$ over candidate behaviors) subject to the condition that the sum of the adoption costs of the behaviors are less than the resource constraint. That is $\sum_i c^i \leq r(v)$. At every epoch, the individual v evaluates all behaviors, including behaviors already adopted, to evaluate payoff. The behavior diffusion process continues until no additional adoption is possible.

In our diffusion model, we assume that the total resources available $r(v)$ at each node are known, while the threshold for adoption θ for any behavior is unknown. This assumption is reasonable if when people are willing to make public their available resources to participate in a set of behaviors. This can arise say in a private, mobile social network app focused on adoption of healthy behaviors including wellness, healthy eating and exercise, where individuals join the network to participate in healthy behaviors but each individual is resource limited. An individual may declare that she has only one hour to spend on exercise each week, but would like to be nudged to participate in a health-related activity.

Figure 3 shows an illustration of the dynamics with a four node network where three different behaviors - (1) recycling, (2) using public transport and (3) eating organic food are spreading. At time step t the state of the network is shown in 3b. At this time step, for v, the social signal of eating organic food is weak. So v considers only recycling and using public transport. After maximizing payoff subject to the resource constraint, v adopts only recycling. Although public transport has strong social signal, v cannot adopt that behavior because it does not have enough resource. Notice

that the payoff for recycling is higher than that of public transport, though the intrinsic utility of recycling was lower than that of public transport.

Next we present a simplified version of the aforementioned multiple behavior diffusion model. Although the previous model makes fewer assumptions and is a more faithful representation of reality, it is cumbersome to deal with analytically. The simplified model, which we will call the "Sticky" model, makes stricter assumptions and affords us some interesting analytical insights on the fundamental processes of the multiple behavior diffusion phenomena.

The difference between the simplified model and the original model is in the temporal unfolding of the diffusion process. Unlike the original model, in the simplified model once a node adopts a behavior at some time step, it does not get rid of that behavior in the subsequent time steps. The behavior adoption is *progressive* [Kempe et al. 2003]. So once a behavior is adopted by a node, it "sticks" with that node (hence the name *Sticky* model). At each time step each node selects the candidate behaviors for adoption from only those behaviors that are not already adopted by that node. Otherwise the selection criteria are the same as the original model. The node then adopts the set of candidate behaviors that maximize his payoff subject to the constraint that the combined cost of the newly adopted behaviors is less than the remaining resource of the node.

3.2. Measurement of the Diffusion

We measure the effectiveness of the diffusion process with three metrics: total participation, total adoption and resource utilization. Since the behavior adoption is a stochastic process, we compute the expected value of each metric through simulation.

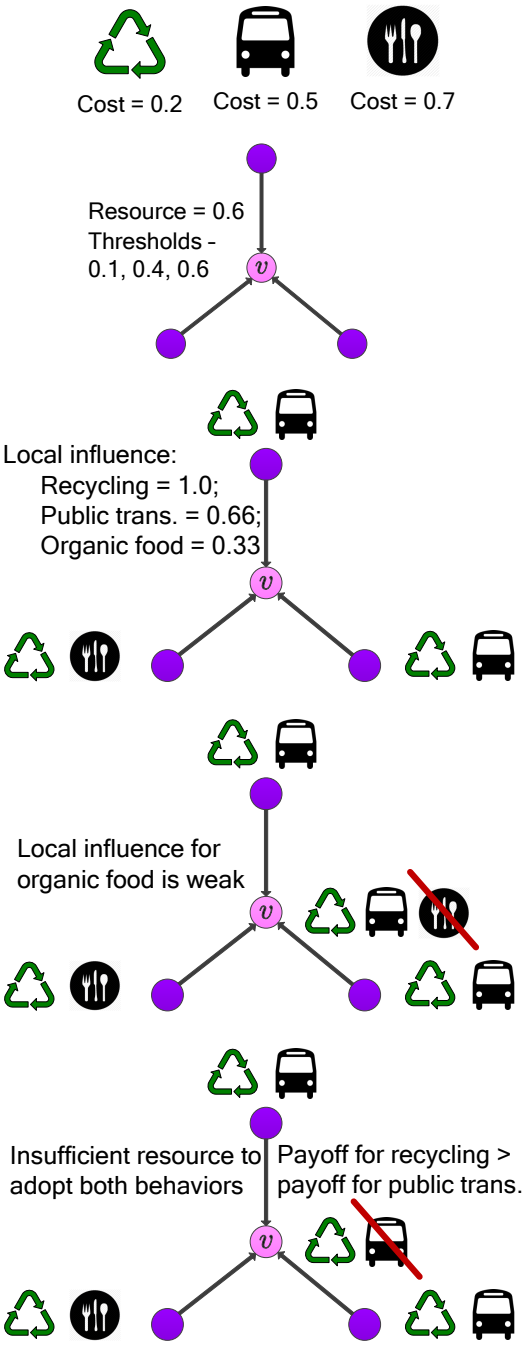
3.2.1. Total Participation. This metric counts the the expected number of individuals who have adopted at least one behavior (i.e. become active) during the process. One goal for an advertiser who is interested in behavior adoption may be to maximize the total number of unique adoptees. Exact computation of this metric is an intractable problem (#P-hard - [Chen et al. 2010]).

3.2.2. Total Adoption. In contrast to total participation, we need to keep track of the total number adoptions of any behavior during the diffusion process. This metric counts the expected number of adoptions over all the behaviors. Notice that since an individual can adopt more than one behavior, total adoption cannot be less than total participation. For the familiar single behavior adoption problem, these two metrics will have same value.

3.2.3. Resource Utilization. This metric captures the *efficiency* of the network to adopt costly behaviors. Not all resources available in a social network may be used for behavior adoption. This is because individuals have variable resources, and they may be unable to adopt the subset of behaviors that fully takes advantage of their desire to participate because of two reasons. First, they may have many more resources than needed to adopt a behavior. Second, if their friends have limited resources, then the social signals that they receive will be about adopting low-cost resources, and hence a particular individual may never see costly behaviors in their social circle that they could potentially adopt. let us assume that a node v with resource $r(v)$ has adopted one or more behaviors. At the end of the diffusion process this individual has adopted a set of behaviors with total cost of adoption s, where $s \leq r(v)$. Therefore the individual has $r(v) - s$ amount of his resource remaining unused. *Resource utilization* is the expected ratio of total utilized resource to the total amount of available resource of all the individuals in the social network.

4. THE SEED SELECTION PROBLEM

In this section we introduce the first algorithmic problem that we want to address in the context of multiple behavior diffusion in a resource constrained social network. This problem is called *seed selection* problem. In the next section we will formally define this problem. Then we will describe and analyze different strategies for solving the problem. Then we will provide experimental evaluation of those strategies on synthetic as well as real world social networks.



(a) *The three beahviors - (1) recycling, (2) using public transport, and (3) eating organic food with respective costs as well as the network is shown. The intrinsic utility of the behaviors are same as the cost. So* $c^1 = u^1 = 0.2$, $c^2 = u^2 = 0.5$, $c^3 = u^3 = 0.7$. *Resource of the node* v, $r(v) = 0.6$, *and the thresholds are -* $\theta^1(v) = 0.1$, $\theta^2(v) = 0.4$, $\theta^3 = 0.6$.

(b) *This is the network at some time step* t. *All three of the neighbors of* v *have adopted recycling, two of them have adopted public transport, and only one of them is eating organic food.* v *has not adopted any behavior yet. The local influences for the three behaviors are as follows -* $l^1(v) = 1.0$, $l^2(v) = 0.66$, $l^3(v) = 0.33$.

(c) *Local influence for organic food is less than the threshold, i.e* $l^3(v) < \theta^3(v)$. *So* v *will not consider organic food for adoption.*

(d) $c^1 + c^2 > r(v)$, *so* v*'s resource is insufficient for adopting both recycling and public transport. Payoff for recycling,* $p^1(v) = 0.6$, *and payoff for public transport,* $p^2(v) = 0.58$. *For* v *the payoff for recycling is higher than the payoff for using public transport though the intrinsic utility of public transport is higher than that of recycling. So* v *will adopt recycling at the end of time step* t.

Fig. 3: *Multiple behavior adoption model.*

4.1. Problem Definition

There are two key problems: we need to identify the set of early adopters or seed nodes and we need to determine which behaviors ought to be adopted by each seed node. We assume that the number of initial adopters is small in comparison to the size of the network. This is reasonable as it corresponds to an advertiser with a finite budget to persuade the seeds to adopt. Here we identify two subproblems which are related to seed identification. To simplify things, in this section we will assume that the behaviors are uniformly distributed over the seed set.

Next, we identify four subproblems; the first two refer to seed identification while the last two are concerned with behavior distribution in the seeds.

4.1.1. P1: Resource Utilization Maximization:. Given a fixed seed budget b and a fixed distribution of behaviors in the seed set, we want to select b nodes in the network such that the resource utilization metric is maximized.

4.1.2. P2: Total Participation (or Adoption) Maximization:. Given a fixed seed budget b and a fixed distribution of behaviors in the seed set, we are interested in finding b nodes in the network that maximize the total participation (or total adoption) in the network.

4.2. Which Nodes Do We Pick?

In this section we develop algorithms and heuristics to pick seed nodes to address the seed selection problems. First we show that the seed selection problem is NP-complete. Next we show that it is possible to construct an approximation algorithm with good approximation guarantee for the seed selection problem for the Sticky model. Then, we will develop heuristics based on node degree, Influence Weight and Immediate Adoption for the general behavior diffusion model under resource constraint.

4.3. NP-Completeness

It can be easily shown that the optimization problems P1 and P2 are NP-complete. We show that influence maximization problem for LT model, which is proven to be an NP-complete problem [Kempe et al. 2003], is a special case of P1. Let the number of behaviors $k = 1$ and the cost of adoption of that behavior is also 1. Each node v is allocated resource $r(v) = 1$. For these values of the parameters our multiple behavior diffusion model reduces to the LT model of influence propagation and resource utilization can be calculated as the ratio of the spread and total number of nodes in the network. So maximizing the resource utilization translates into maximizing the spread. Same transformation applies to problem P2 also since total participation (and total adoption) is identical to the spread in the one behavior case. Next, we propose a number of heuristics to solve the problem.

4.4. Approximation Algorithm for the Sticky Model

In this section we will provide an approximation algorithm for the problem P2 (total participation maximization) under the simplified Sticky model of multiple behavior diffusion. Note that all the three problems - P1, P2 and P3 are still NP-hard in the Sticky model. We will achieve our goal by first showing that the total participation function satisfies a property called submodularity which will lead us to the construction of a standard approximation algorithm for the problem.

4.4.1. Submodularity of Total Participation. In this section we will show that the total participation function for the sticky model is submodular. Throughout this section we will assume that the k behaviors are indexed in the ascending sorted order of their cost. For a given set S of seeds for the k behaviors, let $\sigma(S)$ denote the *total participation* i.e. the expected number of active nodes at the end of the process. We will show that $\sigma(S)$ is a submodular function.

THEOREM 4.1. *For an arbitrary instance of the Sticky Multiple Behavior Diffusion model the total participation function $\sigma(.)$ is submodular.*

PROOF. The proof consists of two steps - first we define an equivalent alternative process of the Sticky Multiple Behavior Diffusion process and then we prove the submodularity for the equivalent process.

Alternative process: For the alternative model we would like to distinguish between the behaviors a node can adopt and those that it cannot under any circumstances. If the cost of adoption of a behavior is greater than the available resource of a node then that node can never adopt that behavior. Since the behaviors are indexed in the ascending sorted order of their cost, this distinction can be made by a single indicator variable. For each node v, let us define $\kappa(v)$ as the largest index j such that $c^j \leq r(v)$. If v does not have enough resource to adopt any of the behaviors then $\kappa(v) := 0$. At the beginning of the process each node v selects at most $\kappa(v)$ edges by repeating the following random edge selection process $\kappa(v)$ times - v selects the edge vw with probability $b_{v,w}$ and no edge with probability $1-\sum_{w\in N(v)} b_{v,w}$, where $b_{v,w}$ denotes the influence weight exerted by the neighbor w on v and $\sum_{w\in N(v)} b_{v,w} \leq 1$; if an edge is selected then that edge is designated as the *live* edge for the behavior i. All the other edges are considered *blocked* for that behavior. So v selects at most one edge for each of the $\kappa(v)$ behaviors. In this model we start with k sets of seeds for each of the k behaviors. In time step t, a node considers behavior i for adoption if the behavior is not already adopted by it and it is reachable via its live edge designated for that behavior from a node which has already adopted behavior i in time step $t-1$. The node then adopts a subset of all considered behaviors that maximizes its total payoff subject to the constraint that the combined cost is less than the remaining resource. Once a node adopts a behavior it becomes *active*. This model is also sticky in nature. When a node adopts a behavior it never gets rid of it. We will show that this process is stochastically equivalent to the Sticky Multiple Behavior Diffusion model described before.

LEMMA 4.2. *For a given seed set S, the following two distributions over the set of nodes are the same:*

(*1*) *The distribution over active sets obtained by running the Sticky Multiple Behavior Diffusion model to completion starting with S.*
(*2*) *The distribution over sets of active nodes reachable from S via live edges under the random selection of edge model described above.*

First we provide the proof for the simpler case when $k = 1$, i.e. there is only one behavior. This case is very similar to the Linear Threshold (LT) model described in [Kempe et al. 2003]. We only need to consider the nodes v with $r(v) \geq c^1$. If we delete all the nodes with $r(v) < c^1$ (and the associated edges) then our model degenerates to the LT model. Here we repeat that proof. We argue by induction over the time step t. Let S_t be the set of nodes who have adopted behavior 1 at the end of time step t for the Sticky Multiple Behavior Diffusion Model with $k = 1$. We need to know the probability that a node v with $r(v) \geq c^1$ that have not yet adopted behavior 1 at the end of time step t will adopt the behavior in the next time step $t+1$. This probability is the same as the probability that the nodes in $S_t \setminus S_{t-1}$ will push the influence weight of v over its threshold, given that the threshold was not already crossed. This probability is given by $\frac{\sum_{w\in S_t\setminus S_{t-1}} b_{v,w}}{1-\sum_{w\in S_{t-1}} b_{v,w}}$.

For the alternative random model each node v with $r(v) \geq c^1$ selects at most one live edge randomly at the beginning of the process. Under this model we need to compute the probability that a node v with $r(v) \geq c^1$ that has not adopted behavior 1 at the end of time step t will adopt it in the next time step. This probability is precisely same as the probability that the live edge of v comes from one of the nodes in $S_t \setminus S_{t-1}$, given that it did not came from S_{t-1}. This probability is also given by $\frac{\sum_{w\in S_t\setminus S_{t-1}} b_{v,w}}{1-\sum_{w\in S_{t-1}} b_{v,w}}$. So by induction we find that the two processes define the same distribution over the active sets. Next we provide the proof for the general case.

PROOF. We prove the claim by induction on the time step t. Clearly the claim is true for $t = 0$. We define S_t^i as the set of active nodes with behavior i at the end of time step t of the Sticky Multiple

Behavior Diffusion model. Let $S_t := \cup_{i=1}^{k} S_t^i$. Notice that $S_0 = S$. Suppose a node v is not active at the end of time step t and $\kappa(v) = \kappa \neq 0$. Then the probability that v will become active at the end of time step $t+1$ is equal to the probability that the nodes in $S_t \setminus S_{t-1}$ will push the influence weight of at least one of the first κ behaviors over its corresponding threshold value, given that none of those thresholds were already crossed. This probability is

$$1 - \prod_{i=1}^{\kappa} \left(1 - \frac{\sum_{w \in S_t^i \setminus S_{t-1}^i} b_{v,w}}{1 - \sum_{w \in S_{t-1}^i} b_{v,w}} \right)$$

On the other hand we run the live edge reachability process as described above and denote by S_t^i the set of all nodes with behavior i at the end of time step t. Let $S_t := \cup_{i=1}^{k} S_t^i$. If node v is not active at the end of time step t with $\kappa(v) = \kappa \neq 0$, then the probability that it will be active at the end of time step $t+1$ is equal to the probability that at least one of its κ live edges comes from the nodes of $S_t \setminus S_{t-1}$ (with the corresponding behavior), given that none of those live edges came from S_{t-1}. This probability is also given by -

$$1 - \prod_{i=1}^{\kappa} \left(1 - \frac{\sum_{w \in S_t^i \setminus S_{t-1}^i} b_{v,w}}{1 - \sum_{w \in S_{t-1}^i} b_{v,w}} \right)$$

By induction over the time step of the process we see that the distribution over the active sets at the end of the Sticky Multiple Behavior Diffusion process is same as the distribution produced by the alternative live edge process. □

Similar to the Sticky Multiple Behavior Diffusion model, let us define $\sigma'(S)$ as the expected number of active nodes at the completion of the alternative random process. By the previous lemma $\sigma(S) = \sigma'(S)$, for all seed sets S under same distribution of behaviors. We will show that $\sigma'(.)$, hence $\sigma(.)$ is submodular. Let X be a particular choice of live/blocked edges for all nodes. Let $\sigma'_X(S)$ denote the cardinality of the set of active nodes at the completion of the alternative process. Suppose $R(v, X)$ denote the set of nodes, which has at least one behavior that is same as the ones adopted by v, reachable via corresponding live edges under the choice X. Clearly $\sigma'_X(S) = \cup_{v \in S} R(v, X)$.

First we will show that for a fixed choice X, $\sigma'_X(.)$ is submodular. Let S and T be two sets of nodes such that $S \subseteq T$ and v is any node. Let us consider $\sigma'_X(S \cup \{v\}) - \sigma'_X(S)$. This is the number of nodes that are in $R(v, X)$ but not in $\cup_{u \in S} R(u, X)$. This number is at least as large as the number of nodes in $R(v, X)$ but not in the bigger union $\cup_{u \in T} R(u, X)$. Therefore it follows that $\sigma'_X(S \cup \{v\}) - \sigma'_X(S) \geq \sigma'_X(T \cup \{v\}) - \sigma'_X(T)$.

Finally we have

$$\sigma'(S) = \sum_{\text{outcomes } X} \text{Prob}[X].\sigma'_X(S)$$

Since a non-negative linear combination of submodular functions is also submodular, $\sigma'(.)$ is submodular. This completes our proof.

4.4.2. The Approximation Algorithm. We are interested in obtaining an approximation guarantee for the total participation maximization problem under the Sticky multiple behavior diffusion model. For this type of optimization problems involving submodular functions there is a greedy algorithm that approximates the optimum within a factor or $(1 - 1/e - \epsilon)$, where e is the base of natural logarithm and ϵ is any positive real number ([Nemhauser et al. 1978], [Kempe et al. 2003]). So the approximation algorithm gives a performance guarantee of slightly better than 63%. We modify the basic greedy algorithm to adapt it to the multiple behavior case (Algorithms 1, 2). In algorithm 2, *Estimate-Spread* simulates the multiple behavior diffusion model a large number of times to estimate the value of the local participation.

ALGORITHM 1: Approximation algorithm for the sticky multiple behavior diffusion model

Input: $G := (V, E)$ - the social network, $\mathbf{b}$ - a vector of size k containing number of required seeds for each of the behaviors
Output: $\mathbf{S}$ - a vector of size k containing seed sets of required size for all the behaviors
Let $V' := V$ and $\mathbf{S} := \phi$;
repeat
  for *each behavior* i **do**
    Let (u^i, s^i) :=Core-Greedy$(i, b[i], \mathbf{S}, V')$;
  end
  Let $i_{max} := \arg\max_{i \in \{1,\ldots,k\}} s^i$;
  Let $v := u^{i_{max}}$;
  Set $V' := V' - v$;
  Set $S[i_{max}] := S[i_{max}] + v$ and $b[j] := b[j] - 1$;
  if $r(v) \leq c^{i_{max}}$ **then**
    Set $r(v) := c^{i_{max}}$;
  end
until $\mathbf{b} = \mathbf{0}$;

ALGORITHM 2: Core-Greedy algorithm used in the approximation algorithm for the sticky model

Input: i - the behavior, $b[i]$ - number of seeds required for the ith behavior, $\mathbf{S}$ - the set of already selected seeds for all the behaviors, V' - the remaining population of nodes to choose new seeds from
Output: (u, s) - if $b[i]$ is not zero then a tuple consisting of u - the best choice of seed from the population V' for the ith behavior, given the already selected seedset $\mathbf{S}$, and s - its corresponding spread value (total participation)
if $b[i] = 0$ **then**
  Return ('nobody',0)
end
for $v \in V'$ **do**
  $s(v)$:= Estimate-Spread$(i, \mathbf{S}, v)$;
end
Select $u := \arg\max_v \{s(v) | v \in V'\}$;
Return $(u, s(u))$;

Next we describe seed selection heuristics based on node degree, influence weight and expected immediate adoption for the general behavior adoption model.

4.5. Node Degree

In this section we develop heuristics for the general behavior diffusion model under resource constraint that are based on degree of a node. The social capital of an individual increases with increase in number of acquaintances. While the nature of the connections and the specific structure of the network in which an individual is embedded matters, we can assume the node degree as a first order approximation to the "influence" of an individual. Hence heuristics based on node degree exploit this idea. We first discuss the basic heuristic and present some useful variants.

4.5.1. Naïve:. In this variant we rank the nodes according to their degree and assign them different behaviors. This is a naïve extension of the high degree heuristic for the LT model [Kempe et al. 2003]. We test three variants of this heuristic. In the first variant, *naïve degree with random tie breaking and no top up* (see Algorithm 3) variant each seed node is assigned exactly one randomly chosen behavior only if its resource is sufficient for the cost of adoption of the behavior. In the second variant *naïve degree with random tie breaking and top up* each seed node is always assigned one randomly chosen behavior irrespective of its resource level. If its resource is not sufficient for

ALGORITHM 3: Naïve Degree Based with Random Tie breaking and No Top Up

Input: $G := (V, E)$ - the social network, $\mathbf{b}$ - a vector of size k containing number of required seeds for each of the behaviors
Output: $\mathbf{S}$ - a vector of size k containing seed sets for each of the k behaviors
Let $V' := V$ and $\mathbf{S} := \phi$;
repeat
  Select $v := \arg\max_u \{|N(u)| : u \in V'\}$;
  $V' := V' - v$;
  Select j uniformly at random from the set of behaviors i that still need seeds to be assigned: $\{i : \mathbf{b}[i] \neq 0\}$;
  if $r(v) \geq c^j$ **then**
    Set $\mathbf{S}[j] := \mathbf{S}[j] \cup \{v\}$ and $\mathbf{b}[j] := \mathbf{b}[j] - 1$;
    Designate v as an early adopter for behavior j;
  end
until $\mathbf{b} = \mathbf{0}$;

adoption of the behavior we top up its resource so that it can bear the cost of adoption of the assigned behavior. In the third variant *naïve degree with knapsack tie breaking*, each seed node is assigned all the behaviors that will maximize its utility subject to its resource constraint—each node will solve a knapsack problem to decide which set of behaviors to adopt. Notice that degree based heuristics are optimistic—it is possible that seed neighbors do not have resources to adopt the behavior of the seed.

4.5.2. Neighbors With Sufficient Resource:. This heuristic takes in account both the degree and available resource of the neighbors when selecting the seed nodes. For each behavior i we calculate $d^i(v)$ - the number of neighbors of a node v with sufficient resource for adoption of i (i.e. the number of neighbors u with $r(u) \geq c^i$). Clearly $d^i(v)$ is a better indicator of the suitability of selecting v as a seed for the ith behavior than just the node degree. In the *degree and resource ranked* heuristic (see Algorithm 4) we compute $d^i(v)$ for all the nodes, rank them according to the value of this metric and select the required number of seeds for the ith behavior from the top of the ranking. If a node is selected as a candidate seed for more than one behaviors, we break the tie randomly and top up its resource so that it can adopt the randomly assigned behavior. We repeat the process until the required number of seeds are selected for all the behaviors. Neither of the degree based heuristics provide any estimate of the effectiveness of the seed in terms of adoptions. We address this issue next.

4.6. Influence Weight

We compute the *Influence Weight* to estimate the influence of the seed set on its neighbors. We can compute the Influence Weight for a set of seeds by summing over the Influence Weight of individual seeds. Let u be a neighbor of v. Hence v exerts a social influence of weight $1/|N(u)|$ on u. If v is the only active seed in the network then it will exert an Influence Weight of $1/|N(u)|$ on u. Hence the Influence Weight exerted by v on its neighbors is $\sum_{u \in N(v)} \frac{1}{|N(u)|}$. For the multiple behavior case we will restrict the summation over those neighbors u that have enough resource to adopt behavior i. We call this metric *Influence Weight (IW)* of v for the behavior i and denote it by $e^i(v)$. In the next two sections, we describe two heuristics based on the Influence Weight metric.

4.6.1. Rank Based with Top-Up:. We rank all the nodes based on the value of $e^i(v)$ and choose the required number of seeds for behavior i starting from the highest ranked nodes. We perform the same evaluation for all behaviors. If a node is selected as a candidate seed for more than one behaviors, then one of the behaviors is chosen randomly and assigned to the node. If the node does not have sufficient resource to adopt that behavior then its resource is topped up. The process continues until the required number of seeds are allocated to all the behaviors.

ALGORITHM 4: Degree and Resource Ranked Heuristic

Input: $G := (V, E)$ - the social network, $\mathbf{b}$ - a vector of size k containing number of required seeds for each of the behaviors
Output: $\mathbf{S}$ - a vector of size k containing seed sets of required size for all the behaviors
Let $d^i(v) := 0$ for all $v \in V$ and $i \in \{1, \ldots, k\}$;
for *each* $v \in V$ **do**
  for *each behavior* i **do**
    for *each neighbor* u *of* v **do**
      if $r(u) \geq c^i$ **then**
        $d^i(v) := d^i(v) + 1$;
      end
    end
  end
end
Let $V' := V$ and $\mathbf{S} := \phi$;
repeat
  for *each behavior* i **do**
    Let T^i be the set of top $b[i]$ nodes from V' in the decreasing sorted order of $d^i(v)$;
  end
  Let $T := \cup_{i=1}^{k} T^i$;
  Set $V' := V' \setminus T$;
  for *each node* v *in* T **do**
    Select j uniformly at random from the set of behaviors i with $v \in T^i$;
    if $r(v) \leq c^j$ **then**
      Set $r(v) := c^j$;
    end
    Set $S[j] := S[j] \cup \{v\}$ and $b[j] := b[j] - 1$;
    Designate v as an early adopter for behavior j;
  end
until $\mathbf{b} = \mathbf{0}$;

4.6.2. Hill Climbing:. The hill climbing heuristic selects the seeds incrementally with the objective of maximizing the marginal increase of the Influence Weight. In this case while calculating the Influence Weight of a node we do not consider the nodes that have already been selected as seeds for other behaviors. As with the previous heuristic, if the node does not have sufficient resource then it is topped up so that it can adopt the assigned behavior.

4.7. Expected Immediate Adoption

We define *Expected Immediate Adoption(EIA)* of a seed set $\mathbf{S}$ for behavior i, denoted by $\mathrm{IA}^i(\mathbf{S})$, as the expected number of nodes who will adopt behavior i in the next time step. Notice that the exact computation of *total* number of adoptions at the completion of the behavior diffusion process is #P-hard for the LT model [Chen et al. 2010]. However we can compute the $\mathrm{IA}^i(\mathbf{S})$ exactly for our model since it is the expected number of adoptions after exactly one time step. For each neighbor u of the seed set $\mathbf{S}$ we would run the adoption decision process (which is equivalent to solving a knapsack problem) and compute the probability that it will adopt behavior i in the next time step. See Appendix A for a detailed example of this computation. Since the adoption decision processes at each neighbor u are independent, we can compute $\mathrm{IA}^i(\mathbf{S})$ by summing over this probability for all the neighbors u of the seed set $\mathbf{S}$. Notice that the two-step or the three-step adoption probabilities are much more difficult to compute exactly — we would need to simulate the stochastic process to evaluate these two cases. The simulation will significantly increase the computation cost. Our

ALGORITHM 5: Influence Weight Ranked

Input: $G := (V, E)$ - the social network, $\mathbf{b}$ - a vector of size k containing number of required seeds for each of the behaviors
Output: $\mathbf{S}$ - a vector of size k containing seed sets of required size for all the behaviors
Let $e^i(v) := 1$ for all $v \in V$ and $i \in \{1, \ldots, k\}$;
for *each* $v \in V$ **do**
  for *each behavior* i **do**
    for *each neighbor* u *of* v **do**
      if $r(u) \geq c^i$ **then**
        $e^i(v) := e^i(v) + \frac{1}{|N(u)|}$;
      end
    end
  end
end
Let $V' := V$ and $\mathbf{S} := \boldsymbol{\phi}$;
repeat
  for *each behavior* i **do**
    Let T^i be the set of top $b[i]$ nodes from V' in the decreasing sorted order of $e^i(v)$;
  end
  Let $T := \cup_{i=1}^{k} T^i$;
  Set $V' := V' \setminus T$;
  for *each node* v *in* T **do**
    Select j uniformly at random from the set of behaviors i with $v \in T^i$;
    if $r(v) \leq c^j$ **then**
      Set $r(v) := c^j$;
    end
    Set $S[j] := S[j] \cup \{v\}$ and $b[j] := b[j] - 1$;
    Designate v as an early adopter for behavior j;
  end
until $\mathbf{b} = \mathbf{0}$;

Immediate Adoption based heuristic builds up the seed set incrementally by assigning the behavior to the seed that provides the maximum marginal increase to the Immediate Adoption value.

4.8. Greedy Approximation (KKT)

[Kempe et al. 2003] presents a greedy approximation algorithm with approximation guarantee of 63% for the LT model and single behavior case. In section 4.4 we have shown that a modified version of this algorithm provides us with the same approximation guarantee for the simplified Sticky Model. We apply the same algorithm for seed selection in the general behavior diffusion model under resource constraint. Since we did not prove the approximation guarantee for this algorithm under the general model, we use it as a heuristic. We call this heuristic KKT heuristic after the authors of the original paper. This heuristic is very similar in structure to our Immediate Adoption based heuristic 8. But instead of using Immediate Adoption for selecting seeds, this algorithm uses *Total Participation*. Since exact computation of Total Participation is #P-hard, this value is estimated using simulation. Due to the high computational cost involved in the simulation this algorithm is not scalable to large sized networks.

4.9. Variations on the Theme

There are a number of variations possible in the way we select seeds. The first being whether we *top up* the resource of a selected seed or not. *Topping up* a seed means providing him with additional resource to adopt the behavior in case it didn't already possess enough resource. This corresponds

ALGORITHM 6: Influence Weight Based Hill Climbing

Input: $G := (V, E)$ - the social network, $\mathbf{b}$ - a vector of size k containing number of required seeds for each of the behaviors
Output: $\mathbf{S}$ - a vector of size k containing seed sets of required size for all the behaviors
Let $V' := V$ and $\mathbf{S} := \phi$;
repeat
  for *each behavior* i **do**
    Let $T^i :=$Core-Hill-Climbing$(i, b[i], S[i], V')$;
  end
  Let $T := \cup_{i=1}^{k} T^i$;
  Set $V' := V' \setminus T$;
  for *each node* v *in* T **do**
    Select j uniformly at random from the set of behaviors i with $v \in T^i$;
    if $r(v) \leq c^j$ **then**
      Set $r(v) := c^j$;
    end
    Set $S[j] := S[j] \cup \{v\}$ and $b[j] := b[j] - 1$;
    Designate v as an early adopter for behavior j;
  end
until $\mathbf{b} = \mathbf{0}$;

ALGORITHM 7: Core-Hill-Climbing

Input: i - the behavior, $b[i]$ - number of seeds required for the ith behavior, $S[i]$ - the set of already selected seeds for the ith behavior, V' - the remaining population of nodes to choose new seeds from
Output: T^i - the set of $b[i]$ newly selected seeds
Let $e^i(v) := 1$ for all $v \in V \setminus S[i]$ and $i \in \{1, \ldots, k\}$;
for *each* $v \in V \setminus S[i]$ **do**
  for *each neighbor* u *of* v *s.t.* $u \in V \setminus S[i]$ **do**
    if $r(u) \geq c^i$ **then**
      $e^i(v) := e^i(v) + \frac{1}{|N(u)|}$;
    end
  end
end
Let $T^i := \phi$;
for $j = 1$ *to* $b[i]$ **do**
  Select $u := \arg\max_v \{e^i(v) | v \in V' \setminus T^i\}$ $T^i := T^i \cup \{u\}$;
  for *each neighbor* v *of* u *in* $V' \setminus T^i$ **do**
    $e^i(v) := e^i(v) - \frac{1}{|N(u)|}$;
  end
end

to real life events like providing the early campaigner with free items, gift coupons or other free services like free access to recycling facilities etc. Depending on whether we allow seeds to be topped up or not we have two variations of the seed selection algorithm - *Topped Up* (suffix **T** is added to the name of the algorithm) and *No Top Up* (suffix **NT** is added). In the **NT** version only the nodes with sufficient resource for adopting a behavior are considered as candidates for seed selection. On the other hand in the **T** version all the nodes[2] are considered as possible candidates.

[2]Whose total resource will not exceed the normalization of 1.0 if extra resource is added for the adoption of the behavior

ALGORITHM 8: Incremental Expected Immediate Adoption Based Heuristic

Input: $G := (V, E)$ - the social network, $\mathbf{b}$ - a vector of size k containing number of required seeds for each of the behaviors
Output: $\mathbf{S}$ - a vector of size k containing seed sets of required size for all the behaviors
Let $V' := V$ and $\mathbf{S} := \boldsymbol{\phi}$;
repeat
  for *each behavior* i **do**
    Let (u^i, s^i) :=Find-Next-Seed-IA$(i, b[i], \mathbf{S}, V')$;
  end
  Let $i_{max} := \arg\max_{i \in \{1,\ldots,k\}} s^i$;
  Let $v := u^{i_{max}}$;
  Set $V' := V' - v$;
  Set $S[i_{max}] := S[i_{max}] + v$ and $b[j] := b[j] - 1$;
  if $r(v) \leq c^{i_{max}}$ **then**
    Set $r(v) := c^{i_{max}}$;
  end
until $\mathbf{b} = \mathbf{0}$;

ALGORITHM 9: Find-Next-Seed-IA heuristic; selects the node that gives maximum marginal increase of the Immediate Adoption value

Input: i - the behavior, $b[i]$ - number of seeds required for the ith behavior, $\mathbf{S}$ - the set of already selected seeds for all the behaviors, V' - the remaining population of nodes to choose new seeds from
Output: (u, s) - if $b[i]$ is not zero then a tuple consisting of the best choice of next seed from the population V' for the ith behavior, given the already selected seedset $\mathbf{S}$ and its corresponding Expected Immediate Adoption value
if $b[i] = 0$ **then**
  Return ('nobody',0)
end
for $v \in V'$ **do**
  $s(v)$:= Compute-IA$(i, \mathbf{S}, v)$;
end
Select $u := \arg\max_v \{s(v) | v \in V'\}$;
Return $(u, s(u))$;

Another variation is possible depending on whether a node can be selected as a seed for more than one behaviors or strictly one behavior. In the first case a seed may be assigned more than one behaviors and we suffix **M** to the seed selection algorithm. In the second case a seed is assigned exactly one behavior and we use the suffix **S**. It is easy to see that **S** version can never find a solution that is better than the **M** version for the same type of top up regime.

Combining these two types of variations we can have four different variants of each seed selection algorithm - **S-T**, **S-NT**, **M-T** and **M-NT**. In this paper most of the results are for the **S-T** variant. However in Appendix B we present some results comparing these different variations of the seed selection algorithm and discuss a few consequences.

4.10. Simulation Experiments

In this section we describe different simulation experiments and compare the effectiveness of our proposed heuristics for the seed selection problem. We have implemented the multiple behavior diffusion model described in Section 3.1 and the heuristics discussed in Section 4.2 in the NetLogo Programming environment [Wilensky 1999]. In the following experiments we have assumed that we want to spread three behaviors b_1, b_2, b_3 with costs $c^1 = 0.2$, $c^2 = 0.5$ and $c^3 = 0.7$. We have assumed that behavior utility is proportional to cost. Hence our nominal utility values for the

corresponding behaviors are $u^1 = 0.2$, $u^2 = 0.5$ and $u^3 = 0.7$. Finally, we assume that individuals' resources are independent and identically distributed i.e the resource $r(v)$ is uniformly distributed random variable $U(0, 1)$ for all $v \in V$.

4.10.1. Network Topologies. We have used synthetic networks as well as a large real-world network for our experiments. We synthesize network topologies through three social network generation models: preferential attachment [Barabasi and Albert 1999]; Small-world [Watts and Strogatz 1998] and spatially clustered [Stonedahl and Wilensky 2008]. All the synthetic networks have 500 nodes. In the preferential attachment network each new coming node adds one link to one of the existing nodes according to the in-degree distribution. The small world network formation starts with a regular circular lattice where each node is connected to next two nodes in the circular order. In the rewiring stage each edge is rewired with probability $p = 0.2$. In the spatially clustered network average node degree is set to 10. The three synthetic networks exhibit all the important properties—low effective diameter, power law degree distribution and high clustering—found in real world social networks. The real world data set is the ca-GrQc collaboration network form the SNAP network database [Leskovec et al. 2007a]. It is a collaboration network amongst authors who submitted their papers to the General Relativity and Quantum Cosmology category of e-print arXiv.org database. This network has 5242 nodes and 28980 edges.

The network types are abbreviated in the tables with experimental results as follows: PA (Preferential Attachment); SW (Small World); SC (Spatially Clustered); COLL (the ca-GrQc collaboration network form the SNAP network database).

4.10.2. Empirical Evaluation. In this section we compare the seven seed selection heuristics described in Section 4.2 for different network topologies. For the seed selection experiments, we fix the behavior distribution over the seeds: the behaviors are assumed to be uniformly distributed over the seeds. We use a specific fraction α of the population as seeds. In this experiment, we have used $\alpha = 0.1$. This means that for synthetic networks, we use $b = 51$ seeds[3], and $b = 501$ for the real-world network. All the results discussed in this section are for the **S-T** variant of the algorithm[4].

The eight heuristics are abbreviated in the experimental results tables as follows: H1 (Random); H2 (Naïve Degree—No Top-up); H3 (Naïve Degree—Knapsack); H4 (Naïve Degree—Top-up); H5 (Degree and Resource Ranked), IWR (Influence Weight —Ranked), IWH (Influence Weight —Hill Climbing), EIA (Expected Immediate Adoption).

Table I: *Maximum Possible Resource Utilization of different network types. Each node solves the knapsack problem and selects optimal behaviors. Then, we diffuse the behaviors. We are reporting the equilibrium values under two conditions: we fix the thresholds and vary topology (Network Average); we fix a random topology and vary thresholds (Threshold Average). Notice that the the quantum physics collaborative dataset, we cannot report a network average since the topology is fixed.*

Network	Threshold Average	Network Average
PA	0.71	0.71
SW	0.72	0.72
SC	0.73	0.73
COLL	0.73	N/A

Since seed selection sub-problems P2 and P3 are NP-complete (ref. Section 4.2), determining the maximum possible utilization or total participation in the network for the given value of b under uniform behavior distribution is computationally intractable. However, we can estimate the value of maximum possible utilization in the network if we assume that $b = N$, the case when each network

[3]the number of seeds is a multiple of 3, since we have 3 test behaviors

[4]In Appendix B we present the result of comparison between the different variants

node is a seed. First the nodes in the network adopt the subset of behaviors that maximizes their payoff subject to the resource constraint. Then we let the diffusion process run till the network reaches equilibrium. The expected value of the resource utilization at this point will upper bound of resource utilization in that network and enables comparison with our heuristics. Table I provides the value of this maximum possible utilization for different networks. Notice that for three behaviors with costs $c^1 = 0.2$, $c^2 = 0.5$ and $c^3 = 0.7$, it is straightforward to show that the maximum utilization will be bounded by the value 0.78, assuming that the thresholds are obtained from $U(0, 1)$. The fact that the simulation results are slightly lower that 0.78 is because nodes will "align" with their neighbors over time due to the social influence.

There are two sources of randomness in the synthetic network generation models: behavior adoption thresholds at each individual for each behavior and network topology. Since each aspect is independent of the other, we have conducted two different types of simulations. In the first, we pick an arbitrary topology and vary individual thresholds over the different simulation runs. We term this as *threshold average*. In the second type of simulation, we fix the individual thresholds, obtained from the uniform distribution, and vary the topologies over the simulations. We term this as *network average*. Notice that the real-world dataset—ca-GrQc network—has a fixed topology and hence only one type of randomness: variation of the individual thresholds. We use 5000 independent runs of the diffusion process to obtain stable estimates for both threshold and network types of simulations.

Table II: *Resource Utilization under Threshold / Network Average. Both versions of the Influence Weight , heuristics IWR, IWH give excellent results. The differences between the heuristics for the same type of average are statistically significant.*

Heuristics	PA	SW	SC	COLL
H1	0.12 / 0.14	0.15 / 0.15	0.16 / 0.16	0.14 / -
H2	0.22 / 0.24	0.16 / 0.17	0.16 / 0.17	0.18 / -
H3	0.28 / 0.30	0.17 / 0.17	0.16 / 0.17	0.18 / -
H4	0.32 / 0.33	0.17 / 0.17	0.17 / 0.18	0.19 / -
H5	0.35 / 0.36	0.21 / 0.21	0.18 / 0.19	0.20 / -
IWR	0.37 / 0.38	0.21 / 0.22	0.20 / 0.21	0.28 / -
IWH	0.37 / 0.38	0.22 / 0.22	0.21 / 0.22	0.29 / -
EIA	0.34 / 0.34	0.22/ 0.23	0.21 / 0.22	- / -

Table II shows the estimated resource utilization of different networks for threshold and network average simulations for each of the eight seed selection heuristics. The two Influence Weight heuristics (IWR, IWH) show the highest expected utilization. The differences between the heuristics (IWR, IWH) and the other heuristics are statically significant ($p < 0.01$) for the same type—threshold or network—of simulation. The table also reveals an expected result: the network average and the threshold averages are nearly identical for the same heuristic.

Table III presents the Total Participation and Total Adoption under threshold average condition for all the eight heuristics. The Expected Immediate Adoption based heuristic (EIA) shows the best result. The results remain qualitatively unchanged in the network average case.

We compare the performance of the two Influence Weight based heuristics (IWR, IWH) and the Expected Immediate Adoption based heuristic (EIA) against the greedy approx. algorithm KKT. Due to huge computational cost it is not practical to run the greedy algorithm on the previous networks. So we used PA, SW and SC networks of size 100 with $b = 9$ for the purpose of comparison. Table IV presents the results of the comparison for the total participation metric. Notice that for the PA network H6 and H8 performs even better than the KKT approximation algorithm. This is not surprising since KKT may fail to obtain the optimal solution in isolated cases. In the next section, we discuss how to distribute behaviors over the seeds.

Table III: *Total Participation / Total Adoption under Threshold average as % of the network size. Both versions of the Influence Weight heuristics IWR, IWH and EIA heuristics give excellent results. The differences between the heuristics for the same type of average are statistically significant.*

Heuristics	PA	SW	SC	COLL
H1	14.0 / 14.1	17.7 / 18.0	20.1 / 20.6	17.3 / 17.6
H2	27.2 / 28.3	19.7 / 20.3	20.3 / 20.6	20.2 / 22.4
H3	31.6 / 35.7	19.1 / 21.3	18.6 / 20.7	22.2 / 23.4
H4	37.5 / 38.0	21.3 / 21.9	21.1 / 21.8	22.6 / 23.8
H5	41.3 / 41.6	24.7 / 25.4	22.9 / 23.4	22.5 / 23.3
IWR	45.0 / 45.2	25.0 / 25.3	25.8 / 26.3	34.8 / 35.9
IWH	44.0 / 44.4	25.9 / 26.4	25.7 / 26.3	35.5 / 36.6
EIA	51.3 / 52.1	26.9 / 27.5	27.6 / 28.4	- / -

Table IV: *Total Participation / Total Adoption under different networks as % of the network size. Heuristics IWR, IWH an EIA give results quite close to the approx. algorithm.*

Heuristics	PA	SW	SC
KKT	43.7 / 44.5	26.2 / 26.4	27.3 / 27.3
IWR	43.9 / 44.5	22.9 / 23.6	24.6 / 25.1
IWH	33.3 / 33.4	22.9 / 23.6	23.5 / 24.1
EIA	43.9 / 44.5	23.6 / 24.5	23.6 / 24.2

4.11. Equivalence between the Threshold and Network Average Cases

In table II we have seen that the resource utilization values under threshold and network average conditions are almost identical. In this section we will investigate the relationship between these two type of averages. First we will show an exact relation for the regular networks. This special case will provide us with helpful insights for analyzing the more general cases.

Suppose we have n nodes with fixed resource distribution. Each node will have a fixed in-degree ρ. Each node selects ρ in-neighbors uniformly at random from the rest $n-1$ nodes. We assume that only in-neighbors can exert influence on a node. In the **TA** case the nodes choose the in-neighbors at random at the beginning of the simulation and then at the start of each simulation run select the threshold values uniformly at random from the interval $[0, 1]$ (*threshold average* case). In the **NA** case each node chooses threshold values uniformly at random from the interval $[0, 1]$ at the beginning of the simulation and then at the start of each simulation run it chooses its ρ in-neighbors uniformly at random from the rest of the nodes (*network average* case). Both the processes start with a set S of seeds for each of the k behaviors. The diffusion process unfolds over time according to the Sticky multiple behavior diffusion process. We will show that $\sigma_{TA}(S) = \sigma_{NA}(S)$ by proving the following lemma:

LEMMA 4.3. *For a given seed set S, the following two distributions over the sets of nodes are the same:*

(*1*) *The distribution over the active sets at the completion of the diffusion process in the **TA** case.*
(*2*) *The distribution over the active sets at the completion of the diffusion process in the **NA** case.*

PROOF. We prove the lemma by induction over the time step t. Clearly it is true at $t = 0$. Let S_t^i denote the set of nodes with behavior i at the end of time step t, and $S_t := \cup_i S_t^i$. For the **TA** case, suppose v is a node that has not adopted any behavior at the end of time step t and $\kappa(v) = \kappa \neq 0$. As before, the probability that v will become active at the time step $t+1$, given that it was not active

at the previous time step is -

$$\begin{aligned}
&1-\prod_{i=1}^{\kappa}\left(1-\frac{\sum_{w\in S_t^i\setminus S_{t-1}^i} b_{v,w}}{1-\sum_{w\in S_{t-1}^i} b_{v,w}}\right)\\
=&1-\prod_{i=1}^{\kappa}\left(1-\frac{\sum_{w\in S_t^i\setminus S_{t-1}^i} \frac{1}{\rho}}{1-\sum_{w\in S_{t-1}^i} \frac{1}{\rho}}\right)\\
=&1-\prod_{i=1}^{\kappa}\left(1-\frac{|S_t^i\setminus S_{t-1}^i|}{\rho-|S_{t-1}^i|}\right)
\end{aligned}$$

For the **NA** case, again let v be a node that is not active at time step t with $\kappa(v)=\kappa\neq 0$. The probability that v will become active at time step $t+1$, given that it was not active till the previous time step is given by -

$$\begin{aligned}
&1-\prod_{i=1}^{\kappa}\left(1-\frac{\sum_{w\in S_t^i\setminus S_{t-1}^i} b_{v,w}}{1-\sum_{w\in S_{t-1}^i} b_{v,w}}\right)\\
=&1-\prod_{i=1}^{\kappa}\left(1-\frac{\sum_{w\in S_t^i\setminus S_{t-1}^i} \frac{1}{\rho}}{1-\sum_{w\in S_{t-1}^i} \frac{1}{\rho}}\right)\\
=&1-\prod_{i=1}^{\kappa}\left(1-\frac{|S_t^i\setminus S_{t-1}^i|}{\rho-|S_{t-1}^i|}\right)
\end{aligned}$$

Since the in-degree of every node is same, we get the same probability distribution over the active sets in both the cases. □

Consequently we obtain the result that the expected number of active nodes in both the **TA** and **NA** cases are the same for the networks with constant in-degree. In the general case when the networks do not have a constant degree for every node but the randomization over the network structure preserves a fixed degree distribution (as in the case of Power Law or Spatially Clustered networks) we may obtain similar results. However the probability that a node becomes active in time step $t+1$, given that it was not active till time step t would be calculated for a node v with $\kappa(v)=\kappa\neq 0$ and degree $d\neq 0$. Assuming that the distribution over the values d would be the same at the time step t in both the cases (notice that the distribution over the values κ would be the same for both the cases since the initial distribution of node resources are the same), we will obtain similar results. Our experimental results show that this observations about the Sticky model carries over to the general model. In all of the simulation experiments we observe that the estimations of the expected values of the different metrics (total participation, total adoption, resource utilization etc.) for both the **TA** and **NA** cases are almost identical.

5. HOW DO WE DISTRIBUTE THE BEHAVIORS?

In this section we discuss the behavior distribution problem - i.e. how to distribute the different behaviors over the selected seed set. We will first formally introduce the problem as an optimization problem. Next we will discuss different strategies for distributing the behaviors over the seed set. Then we will compare these different strategies through simulation experiments and discuss the pros and cons of each strategy.

5.1. Behavior Distribution Problem

The behaviors adopted by the set of seed nodes have different implications on the metrics. If all nodes adopted the least costly behavior, for example, we would expect total participation to increase, but low resource utilization. The converse would be true in the case when seed nodes are chosen

in such a way that all adopt the most expensive behavior. However, if we want to strike a balance between different behaviors such that all the behaviors are represented in the population, then we will have to distribute all the behaviors over the seed set according to some ratio. Here we formalize this scenario as an optimization problem.

5.1.1. P3: Determination of Optimum Behavior Distribution in the Seed Set:. Given a fixed seed budget b and a lower bound on the number of adoptions of the lowest cost behavior s_{min} (or a lower bound on the total participation of the lowest cost behavior S_{min}), what is the optimum distribution of behaviors in the seed set and the optimum set of b seeds that will maximize the resource utilization while maintaining expected spread of s_{min} for the lowest cost behavior (or S_{min} for the total participation).

5.2. Behavior Distribution Strategies

It should be noted that in the case of multiple behavior diffusion metrics like resource utilization, total participation and total adoption depends not only on the choice of the seeds but also on the distribution of the different behaviors in the chosen seed set. We test following five different distributions of the behaviors in the seed set. In the *highest cost behavior only* distribution we allocate all the seeds to the highest cost behavior and none to the other behaviors. In the *proportional to cost* distribution the behaviors are distributed over the seeds in the ratio of their costs. *Uniform* distribution divides the seeds equally amongst all the behaviors. In the *Inversely proportional to cost* behavior distribution behaviors are distributed over the seeds in the inverse ratio of their costs. So the highest cost behavior gets the lowest number of seeds and the lowest cost behavior gets the highest number of seeds. Finally, in the *lowest cost behavior only* distribution all the seeds are assigned to the lowest cost behavior and no seeds are given to the other behavior.

5.3. Experimental Evaluation

In this section we investigate the effects of the different behavior distribution heuristics across the initial seed set described in the previous section. For this simulation, we use heuristic IWH, since it is one of the best performing seed selection heuristics. We designate the fraction of seeds to be early adopters to be $\alpha = 0.1$. This means that we have $b = 51$ for the synthetic networks and $b = 501$ for the quantum physics collaboration network. As before, we compute the metrics under the threshold average and the network average simulations.

In the tables in this section, we shall use the following notation: Low (All seeds are assigned Lowest Cost Behavior); Inv. (the seeds are allocated behavior in Inverse proportion to behavior cost; Unif. (the behaviors are distributed Uniformly at random); Prop. (the behaviors are distributed Proportional to behavior cost); High (all seeds are allocated the Highest cost behavior).

Table V: *Resource Utilization under Threshold / Network Average. Among the behavior distribution heuristics, assigning every seed the lowest (highest) cost behavior results in the lowest (highest) utilization. Assigning seeds proportional to cost, works as well as the assigning everyone the highest cost behavior.*

Heuristics	PA	SW	SC	COLL
Low	0.23 / 0.23	0.14 / 0.13	0.15 / 0.14	0.18 / -
Inv.	0.33 / 0.35	0.20 / 0.21	0.20 / 0.21	0.27 / -
Unif.	0.37 / 0.38	0.22 / 0.22	0.21 / 0.22	0.29 / -
Prop.	0.38 / 0.40	0.24 / 0.24	0.22 / 0.23	0.31 / -
High	0.38 / 0.39	0.24 / 0.25	0.24 / 0.23	0.31 / -

Table V shows the resource utilization in different networks for the threshold average and the network average simulations. We see that when each seed is either allocated the same low (high) behavior, the utilization is lowest (highest). This is unsurprising as we should expect high utilization to occur when we have high cost behaviors in the network. In Table VI, we show the difference

Table VI: *Total Participation / Total adoption under Network Average for different behavior distributions over seeds. Seeds are chosen under heuristic H7. Notice that when all the seeds are the same behavior (Low, High), the number of unique participants and adoptions are identical.*

Dist.	PA	SW	SC
Low	291.12 / 291.12	166.26 / 166.26	178.78 / 178.78
Inv.	250.91 / 254.52	146.36 / 150.09	149.61 / 154.58
Unif.	234.00 / 236.46	133.38 / 136.04	132.27 / 135.77
Prop.	209.66 / 210.98	118.65 / 119.98	113.29 / 114.91
High	144.49 / 144.49	93.79 / 93.79	86.01 / 86.01

between the number of unique participants and the total number of behavior adoptions. We have omitted the simulations for the threshold average case, due to space limitations. Those simulations are qualitatively similar to Table VI. Notice that when all the seeds have either the same low (or high) behaviors assigned to all of them, there is unsurprisingly no difference between the total number participants and the total number of unique adoptions. As Table VI shows, change to the behavior distribution over the seeds alters the unique number of participants as well as the total adoption. Therefore the seed distributions need to be chosen with care, the appropriate metric in mind. Both uniform and propositional to cost behavior distribution methods seem to hit a sweet spot between utilization and behavior diversity.

6. ACHIEVING A TARGET BEHAVIOR DISTRIBUTION

In this section we address the problem of achieving a target behavior distribution in the network. Formally the problem is defined as -

6.0.1. P4: Obtaining Desired Distribution of Behaviors:. Given a fixed seed budget b, and a given fixed target distribution of behaviors q, how do we identify the b seeds and the initial distribution of behaviors p such that final distribution p_T at time T of behaviors in the population matches the target distribution q?

In problem P4 we are interested in achieving a target distribution. One heuristic is to assign the target distribution as the starting behavior distribution over the seeds. In the following section we discuss simulation results for this strategy.

6.1. Empirical Evaluation

In this section we present experiments on a specific heuristic to achieve a specific target distribution q. We propose the following heuristic: set the behavior distribution p over the seeds, *to be equal* to the target distribution q.

In this section we investigate the effect of seed set behavior distribution on the final distribution of behaviors in the population of active individuals. In addition to uniform, proportional to cost and inversely proportional to cost, we will test to check if we can achieve the following six target distributions for the three behaviors in our simulations: $1:1:1$; $1:2:3$; $1:3:2$; $2:1:3$; $2:3:1$; $3:1:2$ and $3:2:1$. Notice that the target behavior distribution ratios are different from the behavior costs. As a reminder, the costs are $c^1 = 0.2$, $c^2 = 0.5$ and $c^3 = 0.7$. IWH heuristic will is used for seed selection. We use the KL-divergence between target distribution and the actual behavior diffusion distribution to measure if our heuristic achieves the target distribution q.

Table VII shows that the heuristic of $p_{seeds} = q$ works well, and results in low KL-divergence between the equilibrium behavior distribution and the target q. One can improve on the heuristic by slightly underweighting the low-cost behaviors and slightly overweighting the high cost behaviors. This exploits the results of Table VI, which shows that low-cost behaviors spread much farther than do high-cost behaviors.

Figure 4 shows the distributions for ca-GrQc collaboration network under threshold average type simulation with uniform initial distribution. The KL divergence between the equilibrium behavior distribution and the target q is 0.05. We can see that the lowest cost behavior increases its share and

Table VII: *KL Divergence, under threshold average conditions, between the equilibrium behavior diffusion and the target distribution q when the initial behavior distribution over the seeds p_{seeds} is set to the target distribution q, for the synthetic networks.*

Target Dist. q	PA	SW	SC
1:1:1	0.02	0.03	0.06
1:2:3	0.06	0.03	0.05
1:3:2	0.03	0.03	0.04
2:1:3	0.02	0.03	0.06
2:3:1	0.03	0.02	0.05
3:1:2	0.02	0.03	0.05
3:2:1	0.04	0.02	0.04

the highest cost distribution loses its share in the final distribution while the median cost behavior almost maintains its share. Again the modified heuristic of underweighting low-cost behaviors, and overweighting high-cost behaviors will help decrease the KL divergence.

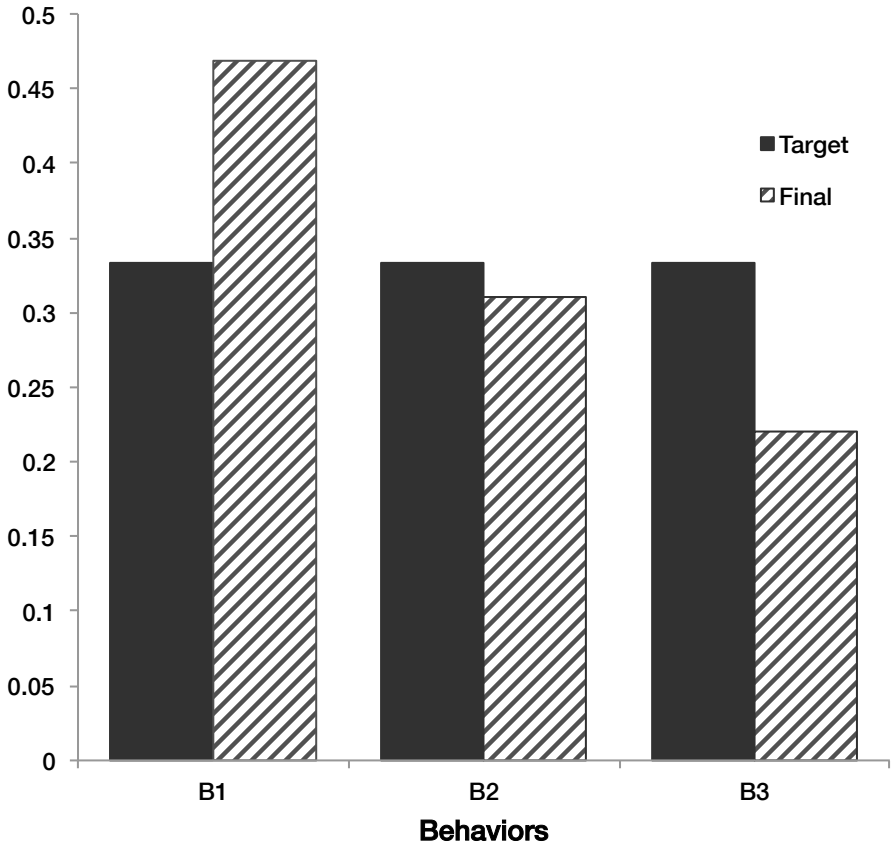


Fig. 4: *A comparison of the final equilibrium distribution and the desired target q of uniform distribution of behaviors when the seed distribution p_{seeds} is set to be uniform. The results are for the collaboration network (COLL). Low cost behaviors spread more than do high-cost behaviors. The KL divergence between the equilibrium behavior distribution and the target q is 0.05.*

Figure 5 shows the comparison in resource utilization with different values of α the fraction of seeds in the network. We show four seed selection heuristics: random, naïve with resource top up, degree and resource ranked seeds and the expected immediate maximization heuristic with hill climbing (H7). In this simulation the behaviors were distributed uniformly across all the seeds. The heuristic H7 consistently outperforms the other three seed selection heuristics, for each value of α.

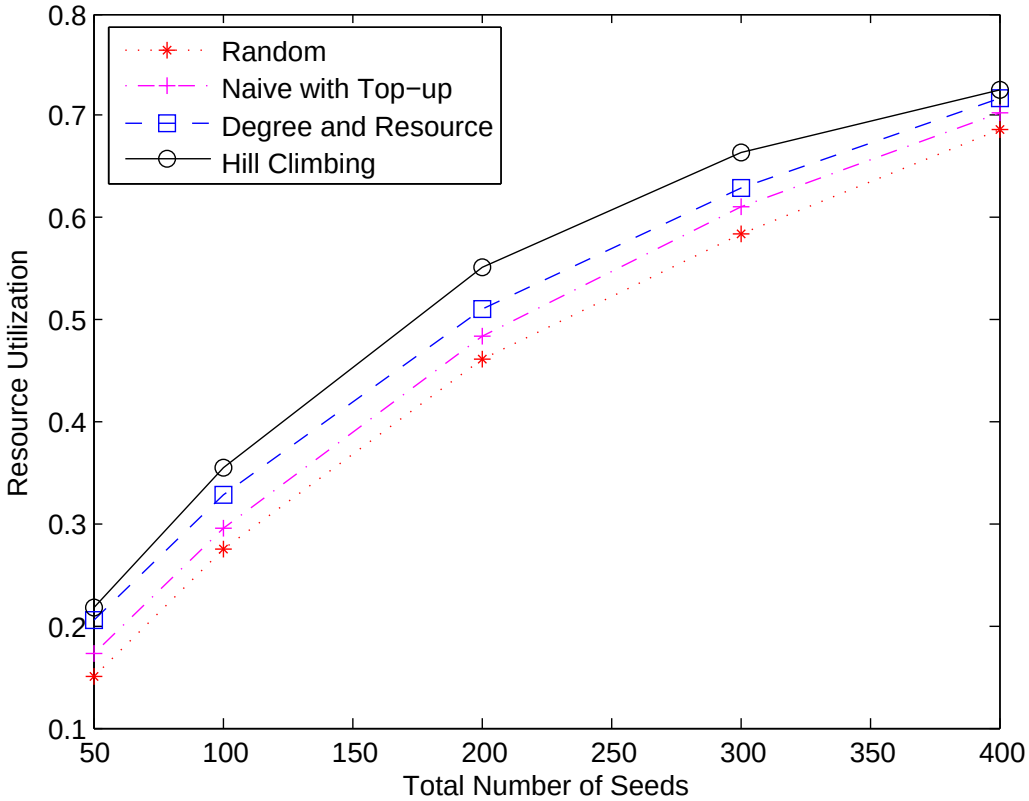


Fig. 5: *The graph shows the comparison in resource utilization with the different number of seeds (that is, different values of the α parameter, for four heuristics. The heuristic H7, which maximizes Influence Weight performs the best. Notice that $l = 50$ corresponds to $\alpha = 0.1$.*

7. DISCUSSION AND OPEN ISSUES

One of the main motivations of the present work was to develop a realistic model of the behavior diffusion process. There are many ways in which our work can be extended. Here we discuss about a few such possible extensions.

Our present model does not consider the role of behavioral inertia in the diffusion process. Often people are hesitant of adopting new behaviors because they cannot free their resources from practicing an old behavior which possibly has less value. This can be modeled in our framework by introducing an additional benefit for the already adopted behaviors. Another technique would be to introduce epidemic models such as SIRS to better model long-term behavioral adoption.

In a network, we receive social signals from our friends, but there is noise because we miss messages and or we check them late. In modeling the behavior adoption problem, we have ignored the role of constraints in how they affect the production and consumption of messages from peers.

Explicit consideration of the cost of social signaling would not only make the model more realistic and provide better bounds on the maximal resource utilization of the networks resources.

8. CONCLUSIONS

In this paper we have considered the problem of seed selection to maximize resource utilization and to achieve a specified target distribution for multiple behavior diffusion processes. We are motivated by collective action problems with applications to sustainability and public health. We have considered a social network where individuals are constrained by available resources for adoption of new behaviors. Our work is the first of its kind, to the best of our knowledge to study the influence of individual resource constraints on multiple, costly behavior adoption. Mindful of the confound between homophily and structural effects, individuals in our model respond to the social influence as well as the intrinsic utility of a spreading behavior. We have shown that the core optimization problems are NP-complete and provided novel heuristics for solving them. We have tested our heuristics against the random and naïve methods and have shown that our heuristics perform very well. We have also shown that assigning the target distribution as the behavior distribution over the seed set results in the final diffused behaviors being very close to the target distribution. Some of the open issues include the use of epidemic models for modeling long-term behavior adoption and incorporating the idea of noisy social signals in modeling behavior adoption.

APPENDIX

A. COMPUTATION OF IMMEDIATE ADOPTION PROBABILITY

In this section we discuss an example of how the different immediate behavior adoption probabilities for a node are computed. This computation depends on whether all the thresholds of a node have the same random value (*matched threshold*) or independent and uniformly distributed random values (*different threshold*). Although all the results in this paper are for the *different threhsold* model, her we present examples for both cases for the sake of completeness. Suppose a vertex v has 8 neighbors. According to our threshold model each of its neighbors exerts an influence of 0.125 on it. Suppose v is already a seed for behavior *A*; moreover it has 2 neighbors with behaviors *B*, and 3 neighbors with behavior *C*. We are interested in computing the probabilities that it will adopt each of the three behaviors in the next time step.

A.1. Matched Threshold:

In this case, for any vertex the thresholds for all the three behaviors will be the same, but it will be assigned independently of other nodes and uniformly at random from the interval $[0, 1]$. So if v's threshold is in the interval $[0, 0.25]$, then v will consider both behaviors B and C together with A for adoption. Our payoff maximizing behavior adoption process dictates that it will adopt a subset of *A,B* and *C* that will provide maximum combined payoff subject to the resource constraint of the node. This adoption decision process is equivalent to solving a knapsack problem. We will solve the knapsack problem and decide which behaviors out of the three behaviors - A, B and C - will be adopted. Any such behavior will be adopted with probability 0.25.

If v's threshold is in the interval $(0.25, 0.375]$ then v will only consider behavior C together with behavior A for adoption. Again after solving knapsack problem and deciding which behaviors to adopt out of A and C, it will adopt any such behavior with probability 0.125.

At last if v's threshold is in the interval $(0.375, 1]$ then it will definitely adopt behavior A - the probability of which is $1 - 0.375 = 0.625$. In the worst case the complexity of this probability computation process for each node is linear order of the number of behaviors.

A.2. Different Threshold:

In the *different threshold* case, for each vertex the thresholds are assigned independently and uniformly at random from the interval $[0, 1]$. So in this case we need to consider all possible combinations of behaviors B and C together with A (which will always be considered) and work out the

individual probabilities. The worst case computational complexity of this process for each node will be exponential in the number of behaviors. In our example we need to consider the following cases:

i) B and C together with A; any behavior selected by the knapsack algorithm will be adopted with probability $0.25 \times 0.375 = 0.09375$.
ii) B together with A; any behavior selected by the knapsack algorithm will be adopted with probability $0.25 \times (1 - 0.375) = 0.15625$.
iii) C together with A; any behavior selected by the knapsack algorithm will be adopted with probability $(1 - 0.25) \times 0.375 = 0.28125$.
iv) Only A; A will be adopted with probability $(1 - 0.25) \times (1 - 0.375) = 0.46875$.

B. VARIANTS OF SEED SELECTION ALGORITHM

Table VIII presents the Total Participation and Total Adoption values for the different variants of the KKT seed selection algorithm and IA based seed selection heuristic. T versions provide better spread than the NT versions which is expected since more resource is required for starting the diffusion in the T version. However for the same type of top up regime there is not much difference between the S and M version. If we consider exact algorithms instead of heuristics and approximation algorithms, then it is easy to see that S version can never produce a result that is better than the M version, since solution for S version is also a valid solution for M version. This fact accounts for the absence of any real difference between the S and T versions in the case of the heuristic and the approximate algorithm.

Table VIII: *Total Participation / Total Adoption under different networks as % of the network size. S and M variants give almost identical results with T variants exceeding NT variants.*

Heuristics	PA	SW	SC
KKT-S-T	43.7 / 44.5	26.2 / 26.4	27.3 / 27.3
H8-S-T	43.9 / 44.5	23.6 / 24.5	23.6 / 24.2
KKT-S-NT	39.5 / 39.5	21.7 / 22.0	22.0 / 22.5
H8-S-NT	39.51 / 39.8	22.7 / 23.2	20.0 / 20.5
KKT-M-T	43.7 / 44.5	26.2 / 26.4	27.1 / 27.1
H8-M-T	39.0 / 45.8	22.8 / 23.5	21.9 / 22.6
KKT-M-NT	39.5 / 39.5	21.7 / 22.0	22.4 / 23.0
H8-M-NT	39.5 / 43.3	22.7 / 23.2	19.7 / 21.1

ACKNOWLEDGMENTS

REFERENCES

Sinan Aral, Lev Muchnik, and Arun Sundararajan. 2009. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences of the United States of America* 106 (2009), 21544.

A.L. Barabasi and R. Albert. 1999. Emergence of Scaling in Random Networks. *Science* 509-12 (1999), 286.

F. M. Bass. 1969. A new product growth for consumer durables. *Management Science* 15 (1969), 215–227.

Shishir Bharathi, David Kempe, and Mahyar Salek. 2007. Competitive influence maximization in social networks. In *In WINE*. 306–311.

Tim Carnes, Rashekhar Nagarajan, Stefan M. Wild, and Anke Van Zuylen. 2007. Maximizing influence in a competitive social network: a followers perspective. In *In ICEC 07: Proceedings of the ninth international conference on Electronic commerce*. ACM, 351–360.

Wei Chen, Yajun Wang, and Siyu Yang. 2009. Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '09)*. ACM, New York, NY, USA, 199–208. DOI: http://dx.doi.org/10.1145/1557019.1557047

Wei Chen, Yifei Yuan, and Li Zhang. 2010. Scalable Influence Maximization in Social Networks under the Linear Threshold Model. In *Proceedings of the 2010 IEEE International Conference on Data Mining (ICDM '10)*. IEEE Computer Society, Washington, DC, USA, 88–97. DOI: http://dx.doi.org/10.1109/ICDM.2010.118

Pedro Domingos and Matt Richardson. 2002. Mining the Network Value of Customers. In *In Proceedings of the Seventh International Conference on Knowledge Discovery and Data Mining*. ACM Press, 57–66.

Sharad Goel, Duncan J. Watts, and Daniel G. Goldstein. 2012. The structure of online diffusion networks. In *Proceedings of the 13th ACM Conference on Electronic Commerce (EC '12)*. ACM, New York, NY, USA, 623–638. DOI: http://dx.doi.org/10.1145/2229012.2229058

Amit Goyal, Francesco Bonchi, and Laks V.S. Lakshmanan. 2010. Learning influence probabilities in social networks. In *Proceedings of the third ACM international conference on Web search and data mining (WSDM '10)*. ACM, New York, NY, USA, 241–250. DOI: http://dx.doi.org/10.1145/1718487.1718518

M. Granovetter. 1983. Threshold Models for Collective Behavior. *Amer. J. Sociology* 6 (1983), 1420–1443.

David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '03)*. ACM, New York, NY, USA, 137–146. DOI: http://dx.doi.org/10.1145/956750.956769

Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. 2007a. Graph evolution: Densification and shrinking diameters. *ACM Trans. Knowl. Discov. Data* 1, 1, Article 2 (March 2007). DOI: http://dx.doi.org/10.1145/1217299.1217301

Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. 2007b. Cost-effective Outbreak Detection in Networks. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. 420–429.

Michael Mathioudakis, Francesco Bonchi, Carlos Castillo, Aristides Gionis, and Antti Ukkonen. 2011. Sparsification of influence networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '11)*. ACM, New York, NY, USA, 529–537. DOI: http://dx.doi.org/10.1145/2020408.2020492

G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. 1978. An analysis of approximations for maximizing submodular set functionsI. *Mathematical Programming* 14, 1 (1 Dec. 1978), 265–294. DOI: http://dx.doi.org/10.1007/bf01588971

Elinor Ostrom, Joanna Burger, Christopher B. Field, Richard B. Norgaard, and David Policansky. 1999. Revisiting the Commons: Local Lessons, Global Challenges. *Science* 284, 5412 (1999), 278–282.

E. Rogers. 1962. *Diffusion of Innovations*. Free Press.

Kazumi Saito, Ryohei Nakano, and Masahiro Kimura. 2008. Prediction of Information Diffusion Probabilities for Independent Cascade Model. In *Proceedings of the 12th international conference on Knowledge-Based Intelligent Information and Engineering Systems, Part III (KES '08)*. Springer-Verlag, Berlin, Heidelberg, 67–75. DOI: http://dx.doi.org/10.1007/978-3-540-85567-5_9

Cosma Rohilla Shalizi and Andrew C. Thomas. 2011. Homophily and Contagion Are Generically Confounded in Observational Social Network Studies. *SOCIOLOGICAL METHODS AND RESEARCH* 40 (2011), 211. doi:10.1177/0049124111404820

F. Stonedahl and U. Wilensky. 2008. NetLogo Virus on a Network model. http://ccl.northwestern.edu/netlogo/models/VirusonaNetwork. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. (2008).

D.J. Watts and S.H. Strogatz. 1998. Collective Dynamics of Small-World Networks. *Nature* 440-42 (1998), 393.

Duncan J. Watts. 2002. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences* 99, 9 (30 April 2002), 5766–5771. DOI: http://dx.doi.org/10.1073/pnas.082090499

U. Wilensky. 1999. NetLogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. (1999).

Online Appendix to: How Do We Find Early Adopters Who Will Guide a Resource Constrained Network Towards a Desired Distribution of Behaviors?

KAUSHIK SARKAR, Arizona State University
HARI SUNDARAM, Arizona State University

DOI : http://dx.doi.org/10.1145/0000000.0000000