# Integrating Social Network Effects in Product Design

Dilek Gunnec, S. Raghavan
Smith School of Business and Institute for Systems Research, University of Maryland, College Park, MD 20742-1815, dgunnec@rhsmith.umd.edu, raghavan@umd.edu

Social network effects on business marketing strategies are becoming inevitable. Taking a step back, we seek to incorporate and analyze such effects on new product development and then propose a model to engineer product diffusion over a social network. We build upon the share-of-choice problem (Kohli and Krishnamurti 1987), which is a combinatorial optimization problem used commonly as one of the methods to analyze conjoint analysis data by marketers in order to identify a product with largest market share, and show how to incorporate social network effects in the share-of-choice problem. We construct a genetic algorithm to solve this computationally challenging (NP-Hard) problem and with simulated data show that ignoring social network effects in the design phase results in a significantly lower market share for a product. In this setting, we introduce the secondary problem of determining the least expensive way of influencing individuals and strengthening product diffusion over a social network. This secondary problem is of independent interest, as it addresses contagion models and the issue of determining influential nodes in a social network, which are of significant interest in marketing and epidemiological settings. Our genetic algorithm obtains near-optimal solutions and is very robust in terms of its running time, scalability, and ability to adapt to additional constraints/variations of the model.

*Key words*: product design; social networks; genetic algorithm; marketing; contagion models; integer programming; optimization; peer influence; share-of-choice

---

## 1. Introduction

Can we isolate our purchase or usage decisions from our societal networks? In this paper, we consider the setting where peer influence plays a significant role in a consumer's choice or there is a tangible benefit from using the same product as the rest of one's social network. Take for example communication tools. Family plans encourage using the same phone network by offering large discounts to customers for in network calls. When a member of a social group has to choose between products, it is natural to take into account the positive network effects (or peer influence) in addition to product's attributes while making a choice. In the extreme case, if the communication is carried over the internet, such as voice or video chat, it is necessary for users to install the same application (e.g., Skype). Interactive information sharing among customers is becoming fast and convenient over the internet with the online social networks or review sections of shopping websites. A friend's (who has similar preferences or concerns) experience with a product can ease the searching process among a variety of product alternatives, especially for complicated technological products. While we are focused on influence over neighborhood relationships, we realize that positive social network influences cause network externalities (Katz and Shapiro 1985).

The exploding reach of the web and the prevalence of social networking sites, which in turn have made large amounts of data on social networks easily available to marketers, has only recently resulted in their recognition as an important tool for marketing (Van den Bulte and Wuyts 2007). Because the market is shifting to the online environment and because of the competitive nature of the industry, it is important for marketing departments to benefit from such information with appropriate marketing strategies. Being able to analyze a social network provides marketers a competitive advantage in terms of forecasting the spread of product influence and intervening at times with promotions or incentives to strengthen this process. Predictions on market share (the number of people who will purchase the product) are critical when a new product enters the market and can create a difference for businesses. Aral and Walker (2011) acknowledge the effects of viral marketing and argue that such viral features can be engineered during the launch of the product. The authors differentiate the "viral characteristics" and "viral features" of a product. The first relates to the content of the product whereas the second corresponds to how the product is shared and how the features allow relationships with the other consumers. Concentrating on the viral features they show that whereas the personalized referrals are more effective in encouraging adoption, passive-broadcast viral messaging is used more often and therefore causes a larger overall adoption. In this paper, we focus more on the viral characteristics of a product implicitly by looking at the changes in the utility consumers get from using the same product with their neighbors. We consider the adoption as a passive-broadcast viral message which increases utilities for the other consumers.

Recent advancements in social media allow better access to social networks of consumers. Innovations such as mobile applications ask the users to log in using their existing Facebook accounts. The online gaming industry is definitely one of the good examples of the modern business model which harvests social network effects among their consumers into their business. Zynga (with a value of more than $15 billion in August 2011) offers a large variety of games which can be played by people over different online social media. The games emulate a real life experience where players are allowed to communicate (send a message or a drink) with each other over an online social network. It is one of the examples of platforms where the users of a product (game in this case) are explicitly connected with each other over a social network which makes it easier to collect information on the relationship network among the customers. In this environment, the product design that best fits the customers can increase the market share significantly.

Product design has been of significant interest both in academic studies and industry applications. Market share forecasts and estimated customer utilities for a product profile allow for a better understanding of the market needs and can lead to better product designs for the companies. Calculation of perceived values for product features have drawn attention from market researchers for many years. Conjoint analysis is one of the most popular tools in new product design for identifying customer preferences and utilities for attribute levels of a product called *part-worth utilities* (Green and Rao 1971). It has been studied widely in the marketing literature and has been used to design many different products in practice. Broadly, there are two main steps in a conjoint analysis. The first is the data collection from consumers and the second is the analysis of this data to obtain part-worth utilities for each customer on each attribute level. After this analysis, part-worth utilities data are used to design the product. Earlier literature on marketing has focused on these two steps, especially in valid data collection and improved statistical estimation, and conjoint optimization to design a product with maximum market share has been given relatively less attention (Camm et al. 2006). In this regard, the share-of-choice (SOC) problem is defined in the product design process where the design corresponds to the selection of levels for each attribute of a product. This is necessary if the number of attributes and levels of the attributes are large and most product designs arising from different attribute level combinations are technologically and economically feasible (Nair et al. 1995). The objective is to create the product profile that will return the largest market share. It finds the best (optimal) design among the different possible product profiles. In this paper, we look at the *combinatorial optimization* problem at the design phase of a product with the same objective, maximizing market share, but we also *explicitly include social network effects* that take place at the product adoption stage, after the product is launched. We aim to integrate the social network effects in the product adoption stage within the share-of-choice problem and propose a genetic algorithm to provide high-quality solutions for this problem.

In the next section, we provide a brief literature review on product design and product influence and define the share-of-choice problem in more detail. Section 3 presents the analytical model that incorporates social network effects into the share-of-choice problem. Additionally, we introduce a complementary problem of identifying the cheapest way of influencing individuals in a social network to achieve the market share associated with a given product design and formulate it as an integer programming model. This problem is of independent interest, as it addresses contagion models and the issue of determining influential nodes in a social network, which are of significant interest in marketing and epidemiological settings. In Section 4, we present a genetic algorithm for the share-of-choice problem with network effects. A novelty of this heuristic procedure is that the fitness (or quality) of a solution is ascertained via an integer program (i.e., by using an exact solution procedure) which allows it to achieve very high-quality solutions. Section 5 provides the results of a computational study demonstrating both the speed of the calculations and the high quality of the solutions provided by the genetic algorithm. In Section 6, several extensions to the share-of-choice problem with network effects are introduced. In particular, we expand the scope of our work to product-line design problem. Section 7 provides concluding remarks.

## 2. Literature Review

Research on consumer behavior shows that consumers' purchase decisions and product evaluations are influenced by their reference groups (Bourne 1957, Burnkrant and Cousineau 1975, Bearden and Etzel 1982, Childers and Rao 1992, Iyengar et al. 2011). Such influence has been termed variously such as bandwagon effect, peer influence, neighborhood effect, conformity and contagion (Iyengar et al. 2009). Although product design is a well-studied subject in marketing and the implications of network externalities on several aspects of marketing (including customer behavior and market structure, product-related decisions such as pre-announcements, timing of product introductions and product differentiation and market entry, see Srinivasan et al. 2004) have been explored, social network effects have generally been ignored within the product design process. In terms of social network effects, the organizational structure and the management of new product development teams have been studied (Leenders et al. 2003, Sosa et al. 2004) with respect to their relation with the design of a product and the creativity involved in the process. To our knowledge, this is the first paper to explicitly incorporate social network (or peer influence) effects among potential customers in product design.

Part-worth utilities are used as inputs for the share-of-choice problem to optimize the selection of levels for each attribute. In this problem, one buys a product only when the utility the person gets by using the product is greater than or equal to her/his "hurdle". The "hurdle" is the utility value at which one would be indifferent between making a purchase or not making a purchase. In this paper we model the social network influence effects on product design by making adjustments to the well studied share-of-choice problem. To state the share-of-choice problem, the following notation is used in the model. Let $V = \{1, 2, \ldots, n\}$ denote the set of people in the market, $h_s$ denote the hurdle utility of person $s \in V$, $K$ be the number of attributes, $L_k$ be the number of levels for attribute $k = 1, 2, \ldots, K$, and $u^s_{kl}$ denote the part-worth utility for person $s$ if level $l$ is chosen for attribute $k$. Data on part-worth utilities are obtained via conjoint analysis studies (Green and Rao 1971). There are two types of binary decision variables; $x_{kl}$ equals 1, if level $l$ has been chosen for attribute $k$ and is 0, otherwise; and $y_s$ equals 1, if person $s$ decides to buy the product and is 0, otherwise. The share-of-choice (SOC) problem is then formulated mathematically as an integer program as follows.

$$
\begin{aligned}
\textbf{SOC:} \qquad & \text{Maximize} \quad \sum_{s=1}^{n} y_s, && (1)\\
& \text{subject to} \quad \sum_{k=1}^{K}\sum_{l=1}^{L_k} u_{kl}^{s} x_{kl} \geq h_s y_s \quad s = 1, 2, \ldots, n, && (2)\\
& \sum_{l=1}^{L_k} x_{kl} = 1 \qquad k = 1, 2, \ldots, K, && (3)\\
& y_s \in \{0, 1\} \quad s = 1, 2, \ldots, n, && (4)\\
& x_{kl} \in \{0, 1\} \quad k = 1, 2, \ldots, K, \quad l = 1, 2, \ldots, L_k. && (5)
\end{aligned}
$$

In this model, the objective is to maximize the market share. Constraint (2) guarantees that people buy the product only if their utilities from the product exceed their hurdles. Constraint (3) ensures that each attribute is assigned only to a single level. Note that no network effects are taken into account in this model.

The share-of-choice problem has been studied in the marketing literature and shown to be an NP-hard problem (Kohli and Krishnamurti 1989). Several heuristics have been used as solution approaches including a divide-and-conquer heuristic (Green and Krieger 1989), greedy search and dynamic programming based heuristics (Kohli and Krishnamurti 1987, 1989), a genetic algorithm (Balakrishnan and Jacob 1996) and a nested partitioning algorithm (Shi et al. 2001). An exact branch-and-bound algorithm (Camm et al. 2006) has been applied more recently for the share-of-choice problem, followed by a branch-and-price approach (Wang et al. 2009) in a product-line setting.[1] In this paper we focus on the design of a single product. Product-line design problems include the design of multiple products. In that setting, there is a constraint of type (2) for each product. As long as the utility from any of the products in the product-line is greater than or equal to the hurdle utility associated with the product, the customer is included in the market share. Although our focus is mainly on the design of a single product, our models and solution methods can be extended to the product-line design problem. However due to the fact that the influence structure can be augmented by the additional influence among users of different product types the model is slightly more complicated than that of a single product design. Hence we discuss these extensions later in Section 6. For a review of methods for product-line optimization in general, the reader is referred to Belloni et al. (2008).

[1] We should note that some commercial packages for conjoint analysis now provide optimization capabilities using some of the heuristics mentioned (Sawtooth Software 2003).

## 3. Incorporating Social Network Information

In this section, we discuss our model to incorporate social network effects in the share-of-choice problem and elucidate why this can be critical in product design. A secondary problem, the least cost influence problem (LCIP), is introduced as a complementary model using a subset of the network representation used for the share-of-choice problem. We also explain why past solution procedures for the share-of-choice problem unfortunately cannot be easily adapted for the share-of-choice problem that incorporates network effects.

### 3.1. Share-of-Choice Model Incorporating Social Network Effects (SOCSNE)

Consistent with social contagion research (Bell and Song 2007, Manchanda et al. 2008), we follow a multiattribute linear utility-maximization approach. We focus on cases where an individual can only observe the outcome of a consumer's purchase decision, but not the relative preferences among each attribute level for three reasons. First, it is less complicated to collect data on purchase decisions among large social network groups using sales data than information about consumers' relative preferences among attributes (in fact due to privacy concerns it may not be feasible to do so). Secondly, potential consumers are exposed to peers' decisions about a product profile over online and offline social networks more frequently since such information sharing requires less proximity or intimacy among consumers. Thirdly, this approach allows for a tractable (and solvable) model for the product design problem while still providing a well-established model to include social network effects. Our mechanism to model social network effects processes the influence from neighbors as an additional attribute of the product. Under this mechanism, the influence from other consumers adds to (or detracts from) product utility of a consumer in a linear additive manner.

In their recent study, Narayan et al. (2011) consider three behavioral mechanisms of how consumers' product choice decisions may be affected from influence of their neighbors. The first is a Bayesian mechanism where a consumer's updated preference for an attribute of the product is a weighted average of her initial (prior) preference and the preferences of their neighbors for the same attribute. The second mechanism is a more generalized Bayesian mechanism and allows for a more flexible process of preference revision. Finally, the third mechanism, which is based on the literature on social contagion and identical to our approach, abstains from updating the relative attribute preferences. Although, they suggest that their first model fits their particular data set (one study of MBA students in the same class) best, Narayan et al. (2011) also find that the mean extent of influence of neighbors' choices on consumer utility is positive and significant. Thus, the choice of a product profile by an influencer leads to an increase in the utility of that profile for the influenced consumer. Narayan et al. (2011) agree that their study might be less representative of the cases where influence among individuals exclude choice-related information sharing and when the number of peers is large. Most importantly, data burdens in the first and the second methods are significant and it is not clear if the required data would be available in a social network setting.

Communication among friends strengthens the inclination towards buying the same product which others have also purchased. The counterpart of this behavior in our model is the decrease in one's hurdle (this could alternatively be viewed as an increase in utility which is the third mechanism in Narayan et al. 2011). The amount of decrease is limited within a hurdle span which we define as the interval between a high and a low hurdle. Hurdle in the traditional product design problem (i.e., where no social network effects are considered) could correspond to a high hurdle. A low hurdle is the smallest value of utility a hurdle can have (i.e., when all friends buy the product). For this paper, motivated by privacy concerns[2] prevalent on social networks, we assume a decrease structure that does not depend on the identity of the neighbor for the influence effect in our model. In particular we use a linearly decreasing influence effect. Note however that *any other influence effect* (nonlinear, increasing etc.) can easily be modeled as described below.

Let $h_s^H$ denote the "high" (H) and $h_s^L$ denote the "low" (L) hurdle for person $s$. The amount of decrease in high hurdle depends on the number of neighbors who purchases the product. We calculate the unit decrease in hurdle for person $s$, $\Delta_s$, as the ratio of the hurdle span and the degree of each node ($deg(s)$): $\Delta_s = \frac{h_s^H - h_s^L}{deg(s)}, \ \forall s \in V$. Using this definition for $\Delta_s$, the share-of-choice model incorporating social network effects can be formulated as using the previous model, SOC, except that constraint (2) is replaced with the following new constraint.

$$\sum_{k=1}^{K} \sum_{l=1}^{L_k} u_{kl}^s x_{kl} \geq h_s^H y_s - \Delta_s \sum_{j \in V} a_{js} y_j \qquad s = 1, 2, \ldots, n. \tag{6}$$

In constraint (6), the right hand side represents the *current* hurdle for person $s \in V$. Each entry of the social network adjacency matrix is shown by $a_{js}$, and it is 1 if there is an edge $(j, s)$ between nodes $j$ and $s$. We refer to the formulation with objective (1), constraints (3), (4), (5) and (6) as the share-of-choice model incorporating social network effects (SOCSNE).

In an environment where the influence effects are negative and have a linear increase structure on the hurdles, $\Delta_s$ would simply be negative in constraint (6). For any other influence structure, the model can be adapted as follows. Let $f_{si}, i, s \in V$, be the amount of decrease in hurdle of person $s$ for the $i^{th}$ additional neighbor who has purchased or adopted the product, and let $g_{si}$ be a binary decision variable which equals 1 if there are at least $i$ neighbors of node $s$ who buy the product,

[2] In 2007, due to protests from users about privacy concerns, Facebook retreated on a tracking program called Beacon which sent messages to users friends about what they are buying on websites like Travelocity.com (Story and Stone 2007). To address this, our model assumes that the anonymity of the users' friends will be preserved and only information on the number of friends purchasing the product is provided. In an environment where such privacy concerns are not important, our model is easily modified as described in Section 6.

and 0 otherwise. Then, constraint (6) would be replaced by constraints (7), (8) and (9). Constraint (7) is similar to constraint (6) in terms of calculating the current hurdle for node $s$. Constraint (8) sets the number of $g_{si}$ variables that are 1 equal to the number of neighbors who adopt the product. Constraint (9) establishes an ordering on the $g_{si}$ variables to correctly compute the change in hurdle.

$$\sum_{k=1}^{K}\sum_{l=1}^{L_k} u_{kl}^{s} x_{kl} \geq h_s^H y_s - \sum_{i=1}^{deg(s)} f_{si} g_{si} \qquad s = 1,2,\ldots,n. \tag{7}$$

$$\sum_{j\in V} a_{js} y_j = \sum_{i=1}^{deg(s)} g_{si} \qquad s = 1,2,\ldots,n. \tag{8}$$

$$g_{s1} \geq g_{s2} \geq \ldots \geq g_{s(deg(s))} \qquad s = 1,2,\ldots,n. \tag{9}$$

### 3.2. Why do Social Network Effects Matter in Product Design?

When social network effects are present, taking them into account in the share-of-choice model can lead to product designs with a higher market share. The following example is constructed to emphasize how large the difference can be among market shares with and without social network effects in the product design process. Consider a simple social network with 3 people where all customers are related to each other, i.e., a fully connected graph with three nodes. The problem is to design a product with a single attribute and maximize the market share. For simplicity, assume that there are only two possible levels for this attribute, level 1 and level 2.

The corresponding data on utility, high hurdle and $\Delta$ values for each person are provided in Table 1. The utilities are $200, 320, 205$ for using level 1; $280, 280, 240$ for using level 2, and the high hurdles are $300, 320, 250$ for the three people, respectively. The $\Delta_s$ values are the same for each person, 20. For each additional neighbor purchasing the product, a person's hurdle decreases by 20. When the original share-of-choice formulation (without the social network effects) is solved, it is easy to see that the optimal solution is to choose level 1; the market share is 1 and only person 2 purchases the product. Although the social network effects have not been taken into account in the design phase, if they are allowed after the product is launched, the purchase by person 2 decreases the hurdles for person 1 and 3. Their current hurdles become 280 and 230 but still are higher than their utilities (insufficient to induce them to buy the product). When the problem is solved taking social networks into account explicitly (SOCSNE), the optimal solution is to choose

**Table 1** **Utility, hurdle and $\Delta_s$ values for the example.**

| Person | Utilities | | High Hurdle | $\Delta_s$ |
|---|---|---|---|---|
| | Level 1 | Level 2 | | |
| 1 | 200 | 280 | 300 | 20 |
| 2 | 320 | 280 | 320 | 20 |
| 3 | 205 | 240 | 250 | 20 |

level 2 and the market share is 3 where all 3 people buy the product. Each hurdle is lowered by 40 in this case because 2 neighbors of each person are purchasing the product (300-40=260 vs. 280, 320-40=280 vs. 280, 250-40=210 vs. 240). In this example, neglecting the social network effects in the product design leads to a solution that is three times worse than the solution obtained by the model that includes social network effects. This example can be generalized (simply add customers to the market identical to person 3) to show that in the worst case, the loss of market share when social network effects are ignored can be as large as the size of the market!

### 3.3. Least Cost Influence Problem (LCIP) over the Buyer Network

Looking at Table 1 in the previous example, the utilities for level 2 for each person are less than their high hurdles which implies no one would buy (adopt) the product at the outset. Yet, the market share equals three when this level is selected for the attribute. Assuming that people are not making buying decisions together, to obtain this market share we need at least one person to purchase the product first and influence others. The secondary problem introduced in this paper concentrates on this ordering of buyers over a finite time period.

In such cases (when each utility is smaller than the corresponding hurdle for each person in the network), to obtain the maximum market share solution from the product design of the SOCSNE model, some people may need to be provided with incentives to make a purchase. In practice, such intervention is not costless, and for a business they can be contemplated as advertising or marketing costs of distributing free samples or discount coupons to a select group of individuals. Here the incentives are paid in units of utilities equal to the difference between one's utility from the product and his/her current hurdle. Once a person makes a purchase, hurdles for her/his neighbors are updated and compared with their utilities to check if they have decreased below them. Every time a new person adopts the product, this update and comparison is repeated. If at some point there is no new buyer and the market share has not yet reached the amount dictated by the SOCSNE model, a new person needs to be given incentives to purchase the product. Although providing incentives result in a larger market share (equal to the amount dictated by the SOCSNE model), it could be expensive. This trade-off is the subject matter of the second problem in this paper.

The objective of the least cost influence problem (LCIP) is to accomplish the market share of the SOCSNE model while minimizing the total amount of incentives given. We formulate it as an integer program. To choose the set of critical people to give incentives to, we analyze the order of buying. We introduce an artificial time dimension, $t = 0, 1, ..., T$ (where $T$ is the number of time periods), to capture the ordering of buyers. The product profile, the market share, and the individuals who will buy the product are inputs for the LCIP model because this problem is solved *after* the solution to the SOCSNE model is obtained. The social network in this problem, $G'$, is a subset of the network in the SOCSNE model and includes only the nodes, $V'$, that adopt the product as a result of the product profile chosen after solving the SOCSNE model (i.e., it includes only those nodes for which $y_s = 1$ in the SOCSNE model solution) and the edges connecting them, $E'$. There are two types of decision variables; $z_s$, $s \in V'$, represents the amount of incentive given to person $s$ and $y_{st}$, $s \in V', t = 0, 1, \ldots, T$, is a binary variable which is 1, if person $s$ buys in period $t$ and is 0, otherwise. Since the attribute levels have been chosen by the first model (SOCSNE), the utility one gets by using the product can simply be represented as one parameter, $U_s$, for each person $s \in V'$. It is the summation of utilities from each level selected in each attribute. The mathematical formulation is as follows:

$$\textbf{LCIP:} \quad \text{Minimize} \quad \sum_{s \in V'} z_s, \tag{10}$$

$$\text{subject to} \quad U_s \geq h_s^H y_{s0} \qquad \forall s \in V', \tag{11}$$

$$U_s \geq h_s^H y_{st} - z_s - \Delta_s \sum_{j \in V'} a'_{js} y_{j,t-1} \qquad \forall s \in V', \forall t \geq 1, \tag{12}$$

$$y_{st} \geq y_{s,t-1} \qquad \forall s \in V', \forall t \geq 1, \tag{13}$$

$$y_{sT} = 1 \qquad \forall s \in V', \tag{14}$$

$$y_{st} \in \{0, 1\} \qquad \forall s \in V', \forall t \geq 0, \tag{15}$$

$$z_s \geq 0 \qquad \forall s \in V'. \tag{16}$$

The objective in this model is to minimize the total amount of incentives given to all the people over a finite number of periods. Constraint (11) identifies the people who buy the product in period 0 without any incentives or influence from neighbors. Constraint (12) is similar to constraint (6) in the SOCSNE model. The right hand side of constraint (12) captures the current hurdle. The current hurdle of a person in any time period is calculated as the remainder after the amount of incentives and the social network effects are subtracted from the high hurdle. Note that in the first model (SOCSNE), the hurdle of a person decreases only with social network effects. In the LCIP model, in addition to the decrease from social network effects, hurdles are also affected by the incentives people receive. The second term in the right hand side of the constraint (12), $z_s$, represents this incentive amount for person $s$. In solving this problem, time is an artificial constraint and thus incentives are not associated with time. Consequently, they can be viewed as being given to the customers at the outset. Since the objective is to minimize the amount of incentives, a person would receive incentives only to cover the difference between the right hand side and the left hand side of constraint (12) (i.e., to be persuaded to buy the product). It is certainly possible to add a time index to $z_s$ and require the model to provide the incentive (i.e., $z_{st}$) in the time period that a person first adopts the product but that unnecessarily increases the number of variables and makes the problem harder to solve. The adjacency matrix in this model represents a subset of the original network since the vertex set is changed. It only contains the buyer network of the individuals that were identified as adopters of the product in the SOCSNE model. Note that once $y_{st} = 1$ for $t = k$, $y_{st} = 1$ for $t \geq k$ by constraint (13) and constraint (14) guarantees that every person on the network buys the product by forcing the last period decision to be 1. By this definition of a time period, there may be more than one buyer in a period. This leads to the following observation.

**Observation 1:** The number of periods is less than or equal to the number of people: $T \leq |V'|$.

Given the solution of the SOCSNE problem, the LCIP aims to identify the influential people (i.e., nodes with $z_s > 0$) whose purchase decisions will strengthen product adoption. In a general setting, the LCIP looks for the nodes to catalyze (or proliferate) a diffusion process. In the literature, the contagion or diffusion over a network is modeled fundamentally using two models, the linear threshold model (Granovetter 1978) and the independent cascade model (Goldenberg et al. 2001). In the linear threshold model, a node is "affected" from diffusion when the sum of the influences from its neighbors exceeds a certain threshold. Our model effectively follows the linear threshold model since a node adopts the product if the sum of the influences from its neighbors and any incentives is greater than the difference, $h_s^H - U_s$. Our goal is to minimize the incentives paid out to these influential nodes. The problem of finding influential nodes in a social network has been studied previously, and is of increasing interest in the literature. The most prominent research in this area is the "influence maximization problem" (Kempe et al. 2003). Here the objective is, for a given parameter $k$, to find a set of $k$-nodes to maximize influence over the network. The $k$ individuals are seeded (i.e., provided the product for free) at the beginning of the diffusion process and the diffusion process takes place (with these $k$-nodes exerting an influence on their neighbors). Our LCIP deviates from this approach. The incentives in our model involve partial inducements tailored for a set of individuals and are not seed products. In this approach, an incentive is used to resolve a knot in the diffusion process that will lead to a larger spread with more influenced nodes at the end. Such catalysation addresses the trade-off of minimizing the amount of incentives given and reaching each individual in the network.

The LCIP is a computationally challenging problem which can take significant amount of time to solve a relatively small problem. In fact for a model where $T = |V'|$, it is not feasible to solve the LCIP model as it renders the computer out of memory in less than three hours for the smallest size data set. In Section 5.4, we provide an iterative approach to solve the LCIP optimally and in much shorter time.

## 4. A Genetic Algorithm for Product Design that Incorporates Social Network Effects

A genetic algorithm (GA) (Holland 1975) is an evolutionary search algorithm that imitates natural selection of species in order to reach near-optimal solutions. We use a GA approach for the SOCSNE model for three reasons. First, the time required to solve the presented integer programming model optimally increases rapidly with the size of the problem (making it a nonviable computational option). Second, due to influence effects, even for the small integer programs state of the art solvers like CPLEX have numerical instability and one is unable to find optimal solutions (specifically CPLEX finds incorrect optimal solutions!) to the problem. Third, specialized approaches to solve the original SOC problem cannot be extended easily to the SOCSNE problem. Our GA generates high quality solutions and is robust in terms of computational time.

For the original share-of-choice problem, Camm et al. (2006) use Lagrangian relaxation with branch-and-bound method to solve the problem exactly. In this method, a search tree is developed, where each level of the search tree corresponds to an attribute of the product. So a node down a path in the tree would have some levels of attributes fixed already. The search tree is pruned using logic-based rules. With these rules, a node is fathomed if the path starting at this node cannot produce a feasible solution superior to one that is already known. This is evaluated by checking whether people's hurdles fall in the range of the minimum and maximum utilities a person may have if that path is followed for the product design. The network relationship among prospective customers in this paper prevents such logical inferences since comparison of utilities from the product and the hurdles include social network effects which depend on the buying status of one's neighbors. The calculation of the objective value at each node would still require the solution of the integer programming model proposed in this paper which would significantly slow down the methodology and actually make it computationally intractable. For the product-line design case, efficient methods are presented and compared by Belloni et al. (2008). Their findings are consistent with the literature in terms of supporting the superiority of genetic algorithm over other methods such as dynamic programming, beam search or a greedy heuristic. While simulated annealing is as successful, they report that it has a running time that is one or two orders of magnitude larger than other methods. Balakrishnan and Jacob (1996) demonstrated the use and advantages of genetic algorithms for solving product design problems. Their study provides a starting point for the genetic algorithm in this paper. We use a similar approach to obtain the product profile with the highest market share. However, our genetic algorithm varies significantly from Balakrishnan and Jacob (1996) in several regards including the fitness evaluation. The outline and the details of the genetic algorithm are given next. In our description, we assume the reader has some familiarity with genetic algorithms. A good introduction to genetic algorithms is the text by Goldberg (1989).

**Outline of the Genetic Algorithm for the SOCSNE Problem**

**Input:** *Parameters:* population size, mutation rate, number of generations.

*Data:* number of attributes, number of levels for each attribute, social network of people, high and low hurdles for each person, utilities for each person.

**Output:** Recommended product design, market share of the chosen product design.

***Step 1*** [GENERATE] Generate an initial population of $q$ product profiles. Set $t=0$.

***Step 2*** [EVALUATION] Calculate fitness of each product profile and let BEST = the profile with the largest fitness.

***Step 3*** [CROSSOVER] Perform single-point crossover operation to generate $q$ offsprings.

***Step 4*** [MUTATION] Perform mutation.

***Step 5*** Calculate fitness of each product profile. If the largest fitness > BEST, update BEST.

***Step 6*** [REDUCTION] Reduce the population to half by choosing the ones with greatest fitness. $t=t+1$.

***Step 7*** If $t <$ number of generations, then go to Step 3.
Else, STOP.

**GENERATE:** Each individual in the population is a product profile and is represented by a binary string, size $\sum_{k=1}^{K} L_k$ where $L_k$ is the number of levels for attribute $k$. An *initial population* is generated randomly by assigning one level for each attribute. For example, if the product has 2 attributes, color and size with 2 and 3 levels as (black, white) and (small, medium, large) respectively, then the product with color white and size small would be represented as (01 100).

**EVALUATION:** After the population is generated, each product profile in the population is evaluated for fitness. *Evaluation* of a product profile corresponds to calculating the market share if that product profile is launched in the market. The exact value of market share is easily determined with the given $x_{kl}$ values for that profile (for example, by solving the integer program SOCSNE model with the $x_{kl}$ values fixed). Methods used to calculate the fitness should be chosen carefully. In an earlier approach, we used an approximate fitness function (where the number of people adopting the product was calculated simply by comparing hurdles with the utilities while hurdles are updated after each purchase) for which the results of the genetic algorithm were significantly worse. Using the exact evaluation via an integer program (for example with the solver CPLEX) corresponds to a hybrid approach called MATHEURISTICS (Maniezzo et al. 2009) marrying mathematical programming with metaheuristic approaches.

**CROSSOVER:** In this step, two offspring product profiles are produced by two parent product profiles. Parents are chosen from a given population with respect to their fitnesses using the roulette-wheel mechanism (Michalewicz 1996). This allows individuals with higher fitnesses to be more likely to get selected as parents. The offspring carry properties of both parents. We use a *single-point crossover* to determine how the heritage is carried to the new generation of individuals. A point is chosen randomly from the points where binary representation of each attribute ends. One of the offspring gets the entries before that point from the first parent and the entries after that from the second parent. The other offspring gets the properties of the attributes after that point from the first parent and the properties of attributes before that point from the second parent. This step is carried out until the number of offspring created is equal to the population size, so the size of the population is doubled at the end of this stage.

**MUTATION:** *Mutation* in product profiles are used to incorporate a different direction in the search process. It corresponds to making a change in the product profile and creating an individual whose properties are not all inherited from the parents. Each individual in the population undergoes this step, however mutation occurs with a predefined mutation rate or probability, $\mu$. In this algorithm, mutation is done by changing the level of one of the attributes to another level. If a profile is subject to mutation, each attribute has an equal chance of being changed. Similarly, all other levels are equally likely to be selected to be the new level. When an individual is mutated, only the mutated version stays in the population. So, at the end of this phase, the size of the population stays the same.

**REDUCTION:** The population size is halved by eliminating the individuals with the least market share. The other half of the profiles with higher market share are carried to the next generation.

**STOPPING CONDITION:** The process is repeated until either there is no significant improvement over multiple generations or a predefined number of iterations is reached.

## 5. Computational Results

In this section, we present an experimental study on the following issues. Performance evaluation of the proposed genetic algorithm, the improvement in market share when network effects are considered in product design process, and how the least cost influence problem can be solved.

### 5.1. Data Generation

Similar to earlier work in the literature (Kohli and Krishnamurti 1987, Balakrishnan and Jacob 1996, Shi et al. 2001, Camm et al. 2006, Wang et al. 2009) on the share-of-choice problem, we generated our own data which included the generation of a social network, part-worth utilities and a hurdle span for each individual. Keeping along similar lines with the data generation methods used in the previous studies (Kohli and Krishnamurti 1987, Nair et al. 1995), part-worth utilities are uniformly generated between 0 and 1 and normalized within individuals. Our modeling approach requires a hurdle span for each individual which is characterized by a high and a low hurdle. To represent a heterogeneous preference behavior among the customers, hurdle spans are selected with the following method; 1000 product profiles are randomly generated and for every individual they are ranked in descending order with respect to their total utilities. High hurdle, which could correspond to total utilities of a status-quo product, is randomly selected from the first 500 product

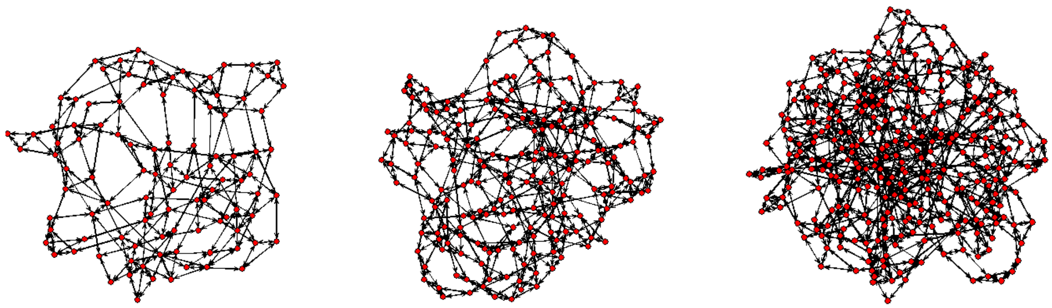

**Figure 1 Social networks of 100, 200 and 300 people for p=0.2, respectively.**

profiles of this ordered set and low hurdle is randomly selected from the second half of the same set.

Social networks of customers are randomly generated. However, the random generation of social networks is more complex than typical random graph generation. The connection topology for social networks lies somewhere between the two extremes of completely regular and completely random (Watts and Strogatz 1998). In our test data, we use the small-world network generation model in R programming (R Development Core Team 2010). In the small-world model every node ends up being only a few (about six) connections away from each other. To be able to generate such a graph, Watts and Strogatz (1998) proposed the following rewiring method. In this method, the network starts with a ring of $n$ vertices, each connected to its $k$ nearest neighbors by undirected edges. A vertex and the edge that connects it to its nearest neighbor in a clockwise sense are chosen. With probability $p$, the edge is reconnected to a vertex chosen uniformly at random over the entire ring, with duplicate edges forbidden; otherwise the edge is left in place. This is repeated until each edge in the original lattice has been considered once. For $p = 0$, the original ring is unchanged; as $p$ increases, the graph becomes increasingly disordered until for $p = 1$, all edges are rewired randomly. Since (i) the clustering coefficient[3] starts to decrease sharply at around $p = 0.3$, and (ii) Watts and Strogatz (1998) show that for intermediate values of $p$ the graph is a small-world network; we used $p = 0.1, 0.2$ and 0.3 as the rewiring probabilities when generating the three social networks with 100, 200 and 300 people, as illustrated in Figure 1.

### 5.2. Performance Evaluation of the GA

To evaluate the performance of the proposed GA, we compare it with the optimal solution of the exact integer programming (IP) model SOCSNE (described in Section 3.1) obtained using CPLEX 12.0 on a 3.40 GHz Pentium 4 processor with 3.49 GB of RAM. The GA was coded in C++ and run on the same computer. The complete combination of product profiles of (3,4,5,6) attributes,

[3] For friendship networks, the intuitive meaning is that it measures the cliquishness of a typical friendship circle (Watts and Strogatz 1998).

(2,3,4,5) levels, (0.1,0.2,0.3) rewiring probabilities ($p$) and (100,200,300) number of people in the market constitute the 144 data sets. After considerable experimentation on small data sets, the GA parameters for the population size and the mutation rate are set to 100 and 0.3, respectively and the stopping condition is set as 10 generations. The market shares for the product profiles selected by the GA and SOCSNE IP model for these 144 data sets are shown in Table 2.

Information on the rewiring probabilities used to generate the social networks and the product profile (in terms of the number of attributes and the number of levels for each attribute) are given in the first three columns of Table 2. Comparing the market shares for the GA and the IP model SOCSNE, the GA obtains the optimal solution in 121 out of 134 data sets where the optimal solution is known (from solving the SOCSNE model). The average running time is 106 seconds for the GA and 1305 seconds for the SOCSNE. The worst case for the SOCSNE IP model is as large as 20110 seconds to find the optimal or 11642 seconds before running out of memory. The running time for the GA is only slightly affected by the size of the social network (the worst case is 199 seconds), whereas the SOCSNE IP model is affected significantly both by the size of the network and the size of the set of possible product profiles, as expected. For 10 problems, the commercial solver was not able to find the optimal solutions because the computer was rendered out of memory (abbreviated as o.o.m.) after (on average) almost 2.4 hours. However the GA found solutions for all of these cases in about 2 minutes.

To explore the GA performance with larger data sets, we ran experiments using product profiles of size (8,9,10) attributes with (2,3,4,5) levels for $p = 0.2$, 100 and 200 people, given in Table 3. The running time of the GA stays around 60 seconds for the first network and 87 seconds for the second network whereas the IP model can take around 15 hours before it reaches the optimal solution. Further it renders the computer out of memory for nine problem sets. On average, for the problems that the IP model could solve, the GA is able to get 98.76% of the optimal solution with the running time around one minute.

### 5.3. Network Effects on Market Share

Two issues are critical in increasing the market share; taking social network effects into account (during or after the product design process) and promoting the product by providing incentives to customers. In Tables 4, 5 and 6, we compare the market shares for product profiles obtained with different solution approaches that consider none, one or both of these two issues. The first solution approach, given in column "SOC", is the exact integer programming model solution of the original SOC problem. In this approach, network effects are not considered (neither in the design process nor afterwards) and incentives are not allowed. People decide to buy the product only if their utilities are greater than or equal to their high hurdles. In the second approach where peer influence

**Table 2 Comparison of market shares for the GA and the optimal solution of SOCSNE.**

| | Num. of attributes | Num.of levels | 100 people GA | 100 people SOCSNE | 200 people GA | 200 people SOCSNE | 300 people GA | 300 people SOCSNE |
|---|---|---|---|---|---|---|---|---|
| p=0.1 | 3 | 2 | 76 | 76 | 132 | 132 | 191 | 191 |
| | 3 | 3 | 76 | 76 | 135 | 135 | 200 | 200 |
| | 3 | 4 | 69 | 69 | 133 | 133 | 199 | 199 |
| | 3 | 5 | 75 | 75 | 147 | 147 | 221 | 221 |
| | 4 | 2 | 76 | 76 | 132 | 132 | 190 | 190 |
| | 4 | 3 | 71 | 71 | 140 | 140 | 185 | 185 |
| | 4 | 4 | 70 | 70 | 133 | 133 | 198 | 198 |
| | 4 | 5 | 80 | 80 | 140 | 140 | 217 | 217 |
| | 5 | 2 | 81 | 81 | 140 | 140 | 206 | 206 |
| | 5 | 3 | 84 | 84 | 146 | 146 | 208 | 208 |
| | 5 | 4 | 78 | 79 | 151 | 151 | 212 | 212 |
| | 5 | 5 | 77 | 77 | 144 | o.o.m. | 208 | 212 |
| | 6 | 2 | 78 | 78 | 136 | 136 | 196 | 196 |
| | 6 | 3 | 77 | 77 | 142 | 142 | 205 | 205 |
| | 6 | 4 | 76 | 80 | 150 | 150 | 218 | 218 |
| | 6 | 5 | 85 | 85 | 145 | o.o.m. | 216 | o.o.m. |
| p=0.2 | 3 | 2 | 68 | 68 | 129 | 129 | 190 | 190 |
| | 3 | 3 | 73 | 73 | 133 | 133 | 191 | 191 |
| | 3 | 4 | 81 | 81 | 129 | 129 | 191 | 191 |
| | 3 | 5 | 74 | 74 | 128 | 128 | 205 | 205 |
| | 4 | 2 | 72 | 72 | 141 | 141 | 191 | 191 |
| | 4 | 3 | 73 | 73 | 136 | 136 | 198 | 198 |
| | 4 | 4 | 77 | 77 | 139 | 139 | 204 | 204 |
| | 4 | 5 | 72 | 73 | 133 | 135 | 198 | 198 |
| | 5 | 2 | 70 | 70 | 127 | 127 | 189 | 189 |
| | 5 | 3 | 82 | 82 | 140 | 140 | 193 | 193 |
| | 5 | 4 | 79 | 79 | 148 | 148 | 203 | 206 |
| | 5 | 5 | 82 | 82 | 144 | o.o.m. | 215 | o.o.m. |
| | 6 | 2 | 78 | 78 | 140 | 140 | 200 | 200 |
| | 6 | 3 | 76 | 76 | 145 | 145 | 206 | 206 |
| | 6 | 4 | 81 | 84 | 146 | 146 | 205 | 209 |
| | 6 | 5 | 80 | 83 | 145 | o.o.m. | 217 | o.o.m. |
| p=0.3 | 3 | 2 | 70 | 70 | 146 | 146 | 196 | 196 |
| | 3 | 3 | 74 | 74 | 134 | 134 | 189 | 189 |
| | 3 | 4 | 82 | 82 | 145 | 146 | 202 | 202 |
| | 3 | 5 | 71 | 71 | 148 | 148 | 199 | 199 |
| | 4 | 2 | 71 | 71 | 142 | 142 | 201 | 201 |
| | 4 | 3 | 66 | 66 | 142 | 142 | 200 | 200 |
| | 4 | 4 | 72 | 72 | 145 | 145 | 202 | 202 |
| | 4 | 5 | 81 | 81 | 141 | 141 | 228 | 228 |
| | 5 | 2 | 72 | 72 | 141 | 141 | 199 | 199 |
| | 5 | 3 | 76 | 76 | 145 | 145 | 208 | 208 |
| | 5 | 4 | 79 | 79 | 145 | 145 | 213 | 213 |
| | 5 | 5 | 76 | 76 | 149 | 151 | 206 | o.o.m. |
| | 6 | 2 | 69 | 69 | 143 | 143 | 208 | 208 |
| | 6 | 3 | 73 | 79 | 143 | 143 | 196 | 196 |
| | 6 | 4 | 76 | 76 | 147 | 147 | 209 | o.o.m. |
| | 6 | 5 | 83 | 83 | 142 | 145 | 215 | o.o.m. |

effects are neglected in the design process but the influence is let to spread among customers after the product is launched in the market, the initial buyers of the product (whose utilities are already greater than their high hurdles) will influence their neighbors, i.e., decrease their hurdles. The updated hurdles for some of these individuals may become less than their utilities changing their decision. They would buy the product resulting in an increase in the market share. This increase in the market share can cascade through the network through neighbors of the new buyers. The market share in this case, where network effects (NE) among customers are taken into account after the product is launched but incentives are not allowed, is given in column "SOC+NE". In the third and the fourth approaches, the product is designed using the GA in which each product profile is evaluated using the exact integer programming solution of the SOCSNE model. However, obtaining this market share may involve promoting the product with incentives (IN), as explained in Section 3.3. For comparison, we present the market shares when incentives are not introduced in column "GA+NE" and the market shares when incentives are used in column "GA+IN" for the same product profile. Note that the market share with the network effects but no incentives is easily computed in an iterative fashion, while the market share with network effects and incentives is computed by using the product profile and inserting it into the IP model SOCSNE (i.e., the variables corresponding to the product are fixed). As should be expected, considering network effects, compared to neglecting these effects (SOC vs. SOC+NE), always increases the market share. Similarly, providing incentives, compared to no incentives (GA+NE vs. GA+IN), produces a larger market share. The computational experiments indicate that this increase in market share can be quite significant, and ignoring network effects leads to a substantially inferior product design. The trade-off between the amount of incentives given and the amount of increase acquired in the market share is analyzed in Section 5.4 with the introduction of a social welfare measure.

### 5.4. Solving the LCIP Model: Identifying Individuals to Pay Out Incentives to

The solution to the least cost influence problem (LCIP) identifies the set of people who receive some incentives to buy the product and further influence their neighbors. The integer programming model for the LCIP was given in Section 3.3 and is computationally intractable to solve even for very small problem sizes. To overcome this problem, we preprocess the model and then use a tractable, iterative, and much faster approach that preserves optimality (i.e., ensures the LCIP is solved to optimality).

In the LCIP model, the influence spreads through the network over a finite number of time periods (Observation 1) and eventually reaches every node. At the beginning of each period, an individual's hurdle is updated with respect to neighbors' buying decisions. Decisions made in time $t$ affect neighbors' hurdles in period $(t+1)$. As a preprocessing before solving the LCIP, we identify by simple comparison the nodes whose utilities are already greater than or equal to their hurdles and the cascade of nodes that purchase the product after being influenced from previous buyers *without* requiring any incentives. Once these nodes are eliminated, the remaining network of nodes are the ones that need an incentive to start a new cascade of buyers (i.e., the nodes that are the difference between GA+IN and GA+NE).

**Table 3** **GA and SOCSNE market share results for larger size data sets.**

| Num. of | Num. of | 100 people | | | | 200 people | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | GA | | SOCSNE | | GA | | SOCSNE | |
| attributes | levels | Market share | Time (sec) | Market share | Time (sec) | Market share | Time (sec) | Market share | Time (sec) |
| 8 | 2 | 68 | 65 | 68 | 2 | 138 | 83 | 138 | 4 |
| 8 | 3 | 87 | 56 | 87 | 122 | 151 | 83 | o.o.m | 7003 |
| 8 | 4 | 82 | 60 | 82 | 23002 | 148 | 83 | o.o.m | 8828 |
| 8 | 5 | 83 | 60 | 84 | 33454 | 155 | 85 | o.o.m | 9651 |
| 9 | 2 | 80 | 61 | 80 | 7 | 156 | 89 | 156 | 11 |
| 9 | 3 | 82 | 59 | 82 | 3795 | 153 | 90 | o.o.m | 2542 |
| 9 | 4 | 84 | 60 | 86 | 15843 | 151 | 92 | o.o.m | 10351 |
| 9 | 5 | 89 | 60 | 91 | 9881 | 156 | 93 | o.o.m | 11920 |
| 10 | 2 | 76 | 64 | 76 | 155 | 137 | 88 | 137 | 27 |
| 10 | 3 | 81 | 62 | 84 | 5443 | 153 | 91 | o.o.m | 8412 |
| 10 | 4 | 80 | 61 | 87 | 55618 | 155 | 87 | o.o.m | 12139 |
| 10 | 5 | 83 | 59 | o.o.m | 16115 | 151 | 88 | o.o.m | 13858 |

**Table 4** **Network effects on market share for rewiring probability p=0.1.**

| Num. of | Num.of | 100 people | | | | 200 people | | | | 300 people | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| attributes | levels | SOC | SOC+NE | GA+NE | GA+IN | SOC | SOC+NE | GA+NE | GA+IN | SOC | SOC+NE | GA+NE | GA+IN |
| 3 | 2 | 35 | 56 | 56 | 76 | 68 | 105 | 102 | 132 | 100 | 154 | 164 | 191 |
| 3 | 3 | 41 | 58 | 61 | 76 | 70 | 115 | 97 | 135 | 100 | 159 | 136 | 200 |
| 3 | 4 | 40 | 56 | 42 | 69 | 58 | 95 | 88 | 133 | 101 | 169 | 169 | 199 |
| 3 | 5 | 40 | 66 | 58 | 75 | 70 | 110 | 123 | 147 | 105 | 184 | 152 | 221 |
| 4 | 2 | 40 | 56 | 66 | 76 | 64 | 111 | 97 | 132 | 96 | 136 | 147 | 190 |
| 4 | 3 | 38 | 55 | 62 | 71 | 68 | 113 | 104 | 140 | 104 | 154 | 144 | 185 |
| 4 | 4 | 37 | 57 | 65 | 70 | 64 | 103 | 107 | 133 | 97 | 171 | 148 | 198 |
| 4 | 5 | 38 | 56 | 61 | 80 | 67 | 104 | 117 | 140 | 96 | 158 | 139 | 217 |
| 5 | 2 | 36 | 51 | 61 | 81 | 59 | 88 | 89 | 140 | 98 | 173 | 168 | 206 |
| 5 | 3 | 40 | 53 | 80 | 84 | 65 | 104 | 88 | 146 | 96 | 149 | 143 | 208 |
| 5 | 4 | 39 | 60 | 66 | 78 | 71 | 116 | 111 | 151 | 95 | 149 | 161 | 212 |
| 5 | 5 | 38 | - | 63 | 77 | 69 | 105 | 117 | 144 | 99 | 160 | 175 | 208 |
| 6 | 2 | 33 | 52 | 57 | 78 | 65 | 96 | 95 | 136 | 91 | 149 | 141 | 196 |
| 6 | 3 | 32 | 47 | 44 | 77 | 69 | 98 | 97 | 142 | 105 | 170 | 179 | 205 |
| 6 | 4 | 41 | 58 | 55 | 76 | o.o.m. | - | 118 | 150 | 104 | 178 | 182 | 218 |
| 6 | 5 | o.o.m. | - | 68 | 85 | o.o.m. | - | 98 | 145 | o.o.m. | - | 168 | 216 |
| Average | | 37.87 | 55.79 | 60.31 | 76.81 | 66.21 | 104.50 | 103.00 | 140.38 | 99.13 | 160.87 | 157.25 | 204.38 |

**Table 5 Network effects on market share for rewiring probability p=0.2.**

| Num. of | Num.of | 100 people | | | | 200 people | | | | 300 people | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| attributes | levels | SOC | SOC+NE | GA+NE | GA+IN | SOC | SOC+NE | GA+NE | GA+IN | SOC | SOC+NE | GA+NE | GA+IN |
| 3 | 2 | 36 | 55 | 42 | 68 | 69 | 84 | 91 | 129 | 103 | 144 | 154 | 190 |
| 3 | 3 | 33 | 48 | 66 | 73 | 71 | 93 | 97 | 133 | 98 | 161 | 138 | 191 |
| 3 | 4 | 34 | 43 | 56 | 81 | 68 | 114 | 110 | 129 | 89 | 133 | 140 | 191 |
| 3 | 5 | 37 | 61 | 53 | 74 | 61 | 88 | 79 | 128 | 103 | 180 | 173 | 205 |
| 4 | 2 | 40 | 55 | 57 | 72 | 65 | 117 | 111 | 141 | 93 | 131 | 154 | 191 |
| 4 | 3 | 36 | 51 | 46 | 73 | 69 | 114 | 113 | 136 | 102 | 152 | 122 | 198 |
| 4 | 4 | 39 | 56 | 48 | 77 | 68 | 110 | 95 | 139 | 93 | 160 | 158 | 204 |
| 4 | 5 | 35 | 59 | 59 | 72 | 64 | 92 | 111 | 133 | 96 | 161 | 140 | 198 |
| 5 | 2 | 37 | 65 | 63 | 70 | 62 | 91 | 87 | 127 | 91 | 144 | 137 | 189 |
| 5 | 3 | 41 | 68 | 75 | 82 | 70 | 103 | 105 | 140 | 97 | 164 | 143 | 193 |
| 5 | 4 | 39 | 59 | 67 | 79 | 67 | 115 | 112 | 148 | 99 | 155 | 174 | 203 |
| 5 | 5 | 37 | 55 | 49 | 82 | 77 | 118 | 131 | 144 | 100 | 174 | 146 | 215 |
| 6 | 2 | 36 | 63 | 72 | 78 | 59 | 91 | 114 | 140 | 97 | 164 | 166 | 200 |
| 6 | 3 | 35 | 61 | 60 | 76 | 70 | 116 | 103 | 145 | 104 | 139 | 153 | 206 |
| 6 | 4 | 37 | 53 | 71 | 81 | 69 | 103 | 113 | 146 | o.o.m. | - | 171 | 205 |
| 6 | 5 | o.o.m. | - | 67 | 80 | o.o.m. | - | 88 | 145 | o.o.m. | - | 164 | 217 |
| Average | | 36.80 | 56.80 | 59.44 | 76.13 | 67.27 | 103.27 | 103.75 | 137.69 | 97.50 | 154.43 | 152.06 | 199.75 |

**Table 6 Network effects on market share for rewiring probability p=0.3.**

| Num. of | Num.of | 100 people | | | | 200 people | | | | 300 people | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| attributes | levels | SOC | SOC+NE | GA+NE | GA+IN | SOC | SOC+NE | GA+NE | GA+IN | SOC | SOC+NE | GA+NE | GA+IN |
| 3 | 2 | 36 | 61 | 61 | 70 | 74 | 112 | 112 | 146 | 109 | 157 | 160 | 196 |
| 3 | 3 | 33 | 50 | 49 | 74 | 67 | 99 | 100 | 134 | 97 | 151 | 136 | 189 |
| 3 | 4 | 36 | 49 | 59 | 82 | 69 | 121 | 103 | 145 | 102 | 155 | 134 | 202 |
| 3 | 5 | 38 | 59 | 46 | 71 | 70 | 121 | 110 | 148 | 97 | 145 | 148 | 199 |
| 4 | 2 | 36 | 67 | 67 | 71 | 69 | 127 | 127 | 142 | 99 | 154 | 141 | 201 |
| 4 | 3 | 39 | 61 | 55 | 66 | 64 | 94 | 119 | 142 | 103 | 168 | 162 | 200 |
| 4 | 4 | 36 | 50 | 56 | 72 | 68 | 98 | 117 | 145 | 99 | 167 | 166 | 202 |
| 4 | 5 | 39 | 55 | 68 | 81 | 72 | 113 | 113 | 141 | 100 | 161 | 161 | 228 |
| 5 | 2 | 36 | 53 | 66 | 72 | 59 | 100 | 105 | 141 | 93 | 157 | 161 | 199 |
| 5 | 3 | 36 | 63 | 50 | 76 | 69 | 106 | 110 | 145 | 102 | 172 | 169 | 208 |
| 5 | 4 | 36 | 52 | 68 | 79 | 72 | 125 | 121 | 145 | 97 | 143 | 167 | 213 |
| 5 | 5 | 38 | 60 | 43 | 76 | 82 | 120 | 124 | 149 | 100 | 154 | 133 | 206 |
| 6 | 2 | 33 | 50 | 64 | 69 | 64 | 106 | 115 | 143 | 86 | 150 | 147 | 208 |
| 6 | 3 | 34 | 57 | 57 | 73 | 73 | 121 | 119 | 143 | 102 | 155 | 161 | 196 |
| 6 | 4 | 35 | 49 | 39 | 76 | 71 | 108 | 102 | 147 | 105 | 163 | 177 | 209 |
| 6 | 5 | o.o.m. | - | 72 | 83 | o.o.m. | - | 130 | 142 | o.o.m. | - | 163 | 215 |
| Average | | 36.07 | 55.73 | 57.50 | 74.44 | 69.53 | 111.40 | 114.19 | 143.63 | 99.40 | 156.80 | 155.38 | 204.44 |

It is then easier to recast the LCIP as follows. For convenience, we will repeat notation and let $V'$ denote the nodes in the problem after preprocessing, $b_s$ denote the difference between the current hurdle and product utility for $s \in V'$ (observe $b_s > 0$ for $s \in V'$), and $d_{js} = a'_{js}\Delta_s$ denote the influence of node $j$ on node $s$ if node $j$ adopts the product. The LCIP can be rewritten as;

$$\begin{aligned}
\textbf{LCIP:}\quad \text{Minimize} \quad & \sum_{s \in V'} z_s, && && (17)\\
\text{subject to} \quad & z_s + \sum_{j \in V'} d_{js} y_{j(t-1)} \geq b_s y_{jt} && \forall s \in V', \forall t \geq 1, && (18)\\
& y_{s0} = 0 && \forall s \in V', && (19)\\
& \text{Constraints } (13), (14), (15) \text{ and } (16).
\end{aligned}$$

Our iterative approach further reduces the size of the LCIP over time periods to obtain an initial solution and increments the number of periods by one at each iteration. We consider the solution of the single period model as the initial solution. In the absence of a successor period, in the single period model network effects cannot take place. The individuals have to be paid the difference between their hurdles and utilities to adopt. In the next iteration, the model has two time periods and network effects are present but limited to only the "first"- and "second"-generation buyers. Note that the cost of total incentives can never be worse in this iteration. In every successive iteration an additional time period is added allowing peer influence effects to cascade further into the network. The iterations are continued until the cost of the LCIP solution (i.e., incentives paid out) stays the same as the cost of the previous iteration. This will be the optimal solution, since it shows that no network effects take place in an additional period and the costs can only decrease when there is at least one new buyer in a period. Table 7 shows the LCIP results for the 32 data sets for $p = 0.2$ and networks of sizes 100 and 200, in particular the ratio between the number of people who decide to buy as a result of their neighbors' decision and who receive incentives, given in column "Return". For example, for the problem set (3,2,100), the return is 0.73 where 15 people are given incentives and they influence 11 additional people to change their decisions without receiving incentives. (Return is calculated as the ratio $\frac{11}{15}$.) Note that the number of new buyers in columns 3 and 7 are equal to the difference between columns "GA+NE" and "GA+IN" in Table 5. This number is the size of the set of buyers who purchase the product after a subset of the network have been given incentives. For the same example, the market share is increased from 42 to 68 by giving incentives to only 15 people. Overall, the returns for the 32 data sets are all greater than or equal to 0.67. The number of iterations needed to solve the LCIP problem are given in Columns 5 and 9 for 100 and 200 people respectively. Notice that the original LCIP model for the problem with 6 attributes, 5 levels, 200 people and 57 new buyers (Table 5) has 145 time periods, but our preprocessing reduces this to a problem over 57 nodes and iterative approach solves the problem within 4 iterations (Table 7).

Although the problem size is decreased with the iterative approach, the running times for the problem data sets still increase with the size of market. The average run time increases from 4.94 seconds to 47 minutes when the number of consumers increase from 100 to 200. There are 3 cases that the computer ran out of memory before reaching the optimal solution. These are the cases for the problem sets of (3,3,200), (3,5,200) and (6,5,200). For these cases, we are still able to get close to optimal solutions in terms of the number of people that are given incentives and the amount of incentives they have been given. This is the case for almost half of the data sets with 300 consumers, so we don't present those results here.

The amount of incentives provided and the increase in the market share are not measured in the same units, perhaps making it somewhat hard to analyze the trade-off. As an alternative, to ease the comparison, we use a welfare measure similar to the utilitarian function (selecting a product on the basis of the sum of utilities) as in Gupta and Kohli (1990). Finding the product profile that results in the largest overall consumer utility is referred to as the "buyer's welfare problem". Individual welfare is calculated as the difference between total utility from a product and the current hurdle of the individual. Social welfare is the sum of all the individual welfares. We compare the two cases; first, the social welfare of the society after they have been given incentives, second, the social welfare if the product spread had no external intervention (no incentives but network effects are included). The impact of incentives over social welfare for 100 and 200 people are given in Table 8. The "Return" column gives the amount of increase in welfare for each unit of incentive provided. Note that all returns are positive and the smallest return is 4 times more than the incentive provided!

### 5.5. An Illustration of the LCIP solution

To further examine and illustrate the cascading network effects in the LCIP solution, we focus on a data set (6,2,200, $p = 0.2$) as an example. For this example, when incentives are not present, the product profile obtained by the GA captures a market share of 114 customers after the network effects (Table 5). However, the product adoption of 9 additional people via incentives increases the market share to 140. The influence originating from these 9 people does not reach the additional 17 people in a single period. Figure 2 and Table 9 show, period-by-period, how the influence spreads over the part of the social network that contains people who will buy the product either after receiving incentives or being influenced by their neighbors. Customers provided with incentives are marked with "+" signs and customers that are influenced by others and buy the product are marked with "•" signs.

**Table 7 LCIP results for the 32 data sets.**

| Num. of attributes | Num. of levels | 100 people: Num. of new buyers | 100 people: Num. of people with incentives | 100 people: Num. of iterations | 100 people: Return | 200 people: Num. of new buyers | 200 people: Num. of people with incentives | 200 people: Num. of iterations | 200 people: Return |
|---|---|---|---|---|---|---|---|---|---|
| 3 | 2 | 26 | 15 | 5 | 0.73 | 38 | 22 | 7 | 0.73 |
| 3 | 3 | 7 | 3 | 3 | 1.33 | 36 | 19 | 7 | 0.89 |
| 3 | 4 | 25 | 13 | 4 | 0.92 | 19 | 7 | 4 | 1.71 |
| 3 | 5 | 21 | 9 | 7 | 1.33 | 49 | 16 | 6 | 2.06 |
| 4 | 2 | 15 | 6 | 2 | 1.50 | 30 | 18 | 6 | 0.67 |
| 4 | 3 | 27 | 13 | 4 | 0.77 | 23 | 13 | 4 | 0.77 |
| 4 | 4 | 29 | 12 | 6 | 1.42 | 44 | 22 | 5 | 1.00 |
| 4 | 5 | 13 | 7 | 3 | 0.86 | 22 | 11 | 5 | 1.00 |
| 5 | 2 | 7 | 3 | 3 | 1.33 | 40 | 17 | 6 | 1.35 |
| 5 | 3 | 7 | 3 | 4 | 1.33 | 35 | 14 | 4 | 1.50 |
| 5 | 4 | 12 | 7 | 3 | 0.71 | 36 | 17 | 6 | 1.12 |
| 5 | 5 | 33 | 15 | 7 | 1.20 | 13 | 4 | 4 | 2.25 |
| 6 | 2 | 6 | 2 | 3 | 2.00 | 26 | 9 | 5 | 1.89 |
| 6 | 3 | 16 | 7 | 3 | 1.29 | 42 | 16 | 8 | 1.63 |
| 6 | 4 | 10 | 6 | 4 | 0.67 | 33 | 16 | 4 | 1.06 |
| 6 | 5 | 13 | 6 | 4 | 1.17 | 57 | 27 | 4 | 1.11 |

**Table 8 Effects of incentives on social welfare.**

| Num. of attributes | Num. of levels | 100 people: Incentives allowed: Social Welfare | 100 people: Incentives allowed: Incentives | 100 people: No Incentives: Social Welfare | 100 people: Return | 200 people: Incentives allowed: Social Welfare | 200 people: Incentives allowed: Incentives | 200 people: No Incentives: Social Welfare | 200 people: Return |
|---|---|---|---|---|---|---|---|---|---|
| 3 | 2 | 26.6057 | 1.7895 | 15.6838 | 6.1032 | 59.6522 | 1.8998 | 44.4102 | 8.0231 |
| 3 | 3 | 45.0922 | 0.3370 | 40.7852 | 12.7817 | 78.5633 | 1.6689 | 58.9479 | 11.7532 |
| 3 | 4 | 38.8263 | 1.1015 | 28.1265 | 9.7139 | 70.0126 | 0.5986 | 58.9194 | 18.5315 |
| 3 | 5 | 44.6026 | 1.0084 | 32.8448 | 11.6597 | 65.7663 | 0.9536 | 40.0679 | 26.9477 |
| 4 | 2 | 43.0862 | 0.4841 | 35.9059 | 14.8316 | 81.7442 | 1.8612 | 67.2795 | 7.7719 |
| 4 | 3 | 38.2466 | 1.2263 | 26.2343 | 9.7952 | 85.6369 | 1.4974 | 75.5858 | 6.7125 |
| 4 | 4 | 44.1616 | 0.7542 | 29.7510 | 19.1081 | 78.1816 | 1.8015 | 56.2452 | 12.1769 |
| 4 | 5 | 48.8871 | 0.7300 | 41.6471 | 9.9178 | 76.8212 | 0.6906 | 65.1930 | 16.8377 |
| 5 | 2 | 48.6230 | 0.6672 | 45.8226 | 4.1975 | 70.9345 | 1.5533 | 52.2856 | 12.0057 |
| 5 | 3 | 59.3606 | 0.6552 | 54.8855 | 6.8302 | 76.8809 | 0.8056 | 59.6078 | 21.4417 |
| 5 | 4 | 56.7058 | 0.7871 | 48.3769 | 10.5820 | 93.9887 | 1.1779 | 72.3549 | 18.3666 |
| 5 | 5 | 55.9945 | 0.9820 | 34.2808 | 22.1120 | 106.4820 | 0.2837 | 97.4306 | 31.8993 |
| 6 | 2 | 50.8347 | 0.2571 | 47.2107 | 14.0970 | 82.6756 | 1.2242 | 67.1765 | 12.6608 |
| 6 | 3 | 43.1903 | 0.5685 | 34.5883 | 15.1297 | 101.5370 | 1.6088 | 76.4503 | 15.5934 |
| 6 | 4 | 53.6025 | 0.5091 | 46.8800 | 13.2057 | 109.6330 | 1.6954 | 89.1506 | 12.0809 |
| 6 | 5 | 65.1313 | 0.5529 | 56.7974 | 15.0719 | 93.1654 | 3.3888 | 64.6902 | 8.4028 |

We focus our attention on the largest network block {9,10,...,16}. In the first period customer 12 is provided with incentives. In the next period, hurdles are updated for its neighbors (11 and 13). Although 11 is influenced by 12, it still does not adopt the product. The high hurdle for 13 decreases sufficiently enough to change its decision. In the third period, the new customers affected from the influence (of node 13) are 11 and 14. Both of them adopt the product. In the following two periods, node 15 buys after the influence from node 14 and node 16 buys as a result of the influence from node 15. In the fifth period node 9 is given incentives to adopt the product. Finally in period 6, the hurdle for node 10 falls below its utilities with the influence from node 9 and it buys the product.

## 6. Extending the SOCSNE Model and Further Research

In this section, we discuss several extensions for the SOCSNE model including the situation when privacy concerns are not a significant issue as well as the product-line design problem. In doing so we also explain how to adapt the GA to address these extensions.

### 6.1. Budget constraint

In many articles that study conjoint analysis, cost has been either explicitly included in the price or has been included in the objective function negatively while maximizing total profit (Dobson and Kalish 1993). In many cases however, because of the financial market conditions, the budget is

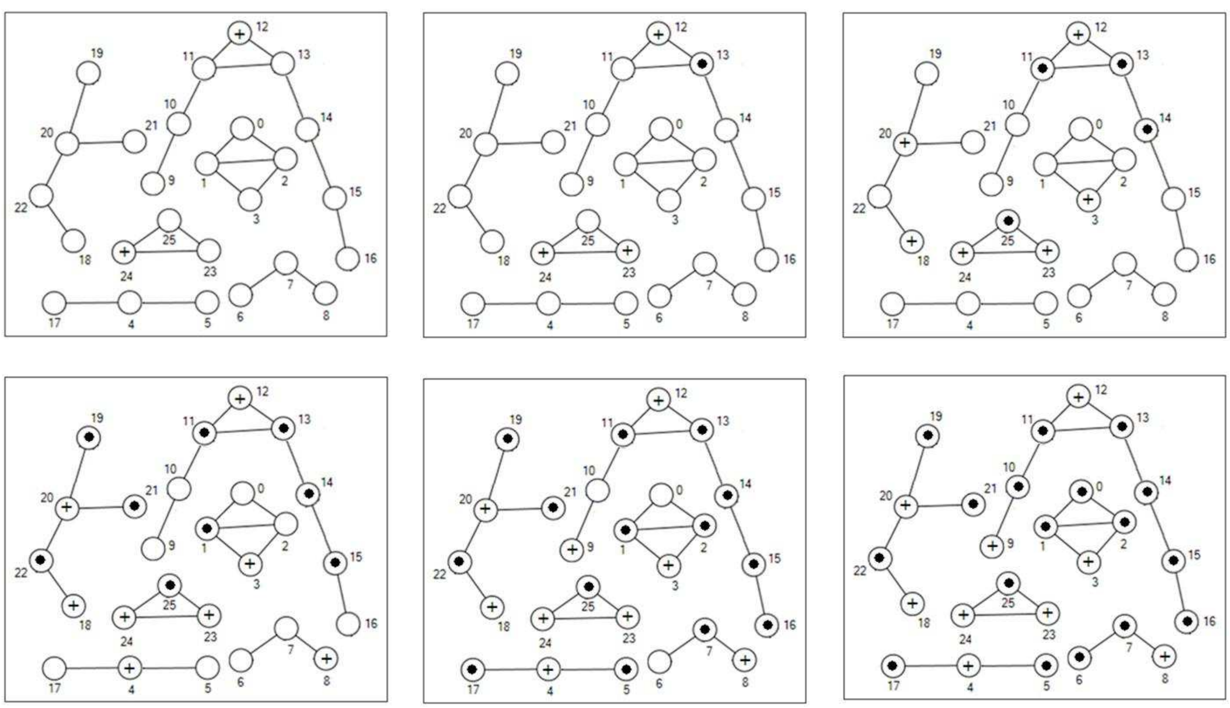

**Figure 2** **LCIP solution for the data set (6,2,200,p=0.2)**

**Table 9 LCIP example.**

| | | | | Current Hurdle | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Person | Utilities | Hurdle | $\Delta$ | t=1 | t=2 | t=3 | t=4 | t=5 | t=6 | Incentives |
| 0 | 2.0899 | 2.5005 | 0.3666 | 2.5005 | 2.5005 | 2.5005 | 2.5005 | 2.1340 | **1.7674** | 0 |
| 1 | 3.5918 | 3.8967 | 0.3801 | 3.8967 | 3.8967 | 3.8967 | **3.5166** | | | 0 |
| 2 | 3.6248 | 3.9235 | 0.2552 | 3.9235 | 3.9235 | 3.9235 | 3.6683 | **3.4131** | | 0 |
| 3 | 3.7329 | 3.8533 | 0.1208 | 3.8533 | 3.8533 | **3.7329** | | | | 0.1204 |
| 4 | 2.0695 | 2.5299 | 0.2614 | 2.5299 | 2.5299 | 2.5299 | **2.0695** | | | 0.4604 |
| 5 | 2.3355 | 2.6251 | 0.3102 | 2.6251 | 2.6251 | 2.6251 | 2.6251 | **2.3148** | | 0 |
| 6 | 2.0638 | 2.2550 | 0.1922 | 2.2550 | 2.2550 | 2.2550 | 2.2550 | 2.2550 | **2.0628** | 0 |
| 7 | 2.4257 | 2.7660 | 0.5966 | 2.7660 | 2.7660 | 2.7660 | 2.7660 | **2.1694** | | 0 |
| 8 | 3.1668 | 3.2157 | 0.2219 | 3.2157 | 3.2157 | 3.2157 | **3.1668** | | | 0.0489 |
| 9 | 2.2630 | 2.3132 | 0.6612 | 2.3132 | 2.3132 | 2.3132 | 2.3132 | **2.2630** | | 0.0502 |
| 10 | 3.5868 | 3.9947 | 0.2445 | 3.9947 | 3.9947 | 3.9947 | 3.7502 | 3.7502 | **3.5057** | 0 |
| 11 | 3.6349 | 4.3801 | 0.3775 | 4.3801 | 4.0026 | **3.6251** | | | | 0 |
| 12 | 2.9853 | 3.0490 | 0.4652 | **2.9853** | | | | | | 0.0637 |
| 13 | 3.7322 | 3.8657 | 0.1374 | 3.8657 | **3.7283** | | | | | 0 |
| 14 | 3.7534 | 4.0310 | 0.5328 | 4.0310 | 4.0310 | **3.4982** | | | | 0 |
| 15 | 3.7941 | 3.9833 | 0.5875 | 3.9833 | 3.9833 | 3.9833 | **3.3958** | | | 0 |
| 16 | 2.0522 | 2.0543 | 0.2552 | 2.0543 | 2.0543 | 2.0543 | 2.0543 | **1.7991** | | 0 |
| 17 | 1.8873 | 2.1816 | 0.4424 | 2.1816 | 2.1816 | 2.1816 | 2.1816 | **1.7391** | | 0 |
| 18 | 5.0710 | 5.1173 | 0.2870 | 5.1173 | 5.1173 | **5.0710** | | | | 0.0462 |
| 19 | 3.4486 | 3.7521 | 0.3519 | 3.7521 | 3.7521 | 3.7521 | **3.4002** | | | 0 |
| 20 | 2.8955 | 2.9618 | 0.1512 | 2.9618 | 2.9618 | **2.8955** | | | | 0.0664 |
| 21 | 2.2459 | 2.6065 | 0.4770 | 2.6065 | 2.6065 | 2.6065 | **2.1295** | | | 0 |
| 22 | 2.5947 | 3.5842 | 0.5028 | 3.5842 | 3.5842 | 3.5842 | **2.5785** | | | 0 |
| 23 | 2.5196 | 3.2482 | 0.4113 | 3.2482 | **2.5196** | | | | | 0.3172 |
| 24 | 2.7059 | 2.7566 | 0.3072 | **2.7059** | | | | | | 0.0507 |
| 25 | 2.8436 | 3.9490 | 0.6422 | 3.9490 | 3.3067 | **2.6645** | | | | 0 |

not always available to develop a product that will return the highest market share. Consequently, it is important to take product profile decisions in the context of budgetary constraints at the manufacturing organization. This is especially true if the costs are in terms of man-power or time, or the focus is a pre-stated design objective (e.g., design a laptop with a manufacturing cost of less than $300 while maximizing market share). Consequently, rather than finding the optimal design that will produce the highest market share, the question is to find the optimal design that will give the largest market share *under a limited design budget*. To address these concerns, we propose adding the following constraint (20) to the SOCSNE model. The design cost associated with level $l$ of attribute $k$ is given by $c_{kl}$ and the total budget is $B$.

$$\sum_{k=1}^{K}\sum_{l=1}^{L_k} c_{kl}x_{kl} \leq B. \tag{20}$$

In this way, the total cost of the selected levels for the attributes will not be allowed to exceed a given design budget. It is very easy to modify the GA for this problem. This constraint can simply be added to the SOCSNE model when calculating the fitness of a product profile within the GA.

### 6.2. Weighted Network Links

In the SOCSNE model, motivated primarily due to privacy considerations, the set of neighbors of $s$ (i.e., all nodes to which $s$ is adjacent) has the same influence on $s$, $\Delta_s$. In a setting where neighbors of a node have different influences (this may be the case when the privacy concerns discussed earlier are moot), we suggest modifying the previous model as follows. Let $\Delta_{js}$ represent the influence of neighbor node $j$ on node $s$ (i.e., the amount by which node $j$'s product adoption reduces node $s$'s hurdle). To calculate $\Delta_{js}$, we propose a weighted directed graph model, where $(j,s)$ represents the directed arc from node $j$ to node $s$ and $a_{js} = 1$ if there is a directed arc from node $j$ to node $s$ with $A(s) = \{j | a_{js} = 1\}$. Further, the relative influence of node $j$ on node $s$ is denoted by the weight $w_{js}$ (analogous to Narayan et al. 2011). For each arc $(j,s)$, $\Delta_{js}$ can now be calculated using the formula;

$$\Delta_{js} = (h_s^H - h_s^L)\frac{w_{js}}{\sum\limits_{j \in A(s)} w_{js}}. \tag{21}$$

If there is no arc from node $j$ to node $s$, then $\Delta_{js}$ is set to 0. After $\Delta_{js}$ is calculated, constraint (6) in the SOCSNE model is replaced with the following inequality.

$$\sum_{k=1}^{K}\sum_{l=1}^{L_k} u_{kl}^s x_{kl} \geq h_s^H y_s - \sum_{j \in V} \Delta_{js} a_{js} y_j \qquad s = 1, 2, \ldots, n. \tag{22}$$

Note that $\Delta_{sj}$ and $\Delta_{js}$ are not necessarily equal in this setting. This is reasonable as influence among people is not necessarily symmetric. Similar to the modification of the GA for the case with the budget constraint, the fitness calculation is easily modified to take into account a SOCSNE model with constraint (22).

### 6.3. Product-line Design

Product-line design problems occur when a company wants to design $M$ products which will appeal to different consumer segments of a single product. As mentioned in Section 2 the product-line design problem with social network effects is more complicated than the single product design problem since the peer influence effects among users of the product would have to be modeled differently. We model this influence in two levels as first and second order peer influence effects. The first order effects are the peer influences from consumers of the same segment, i.e., using the same product, and the second order effects are the peer influence among all product users across the product-line.

Before a purchase decision is made, the utility one gets from a product are dynamic under peer influence, i.e., increases with the neighbors who also buy the product (as the decrease in hurdle is equivalent to an increase in utility). This fluctuation (change) in the utilities due to peer influence prevents one from making a list of preference ordering among the products before solving the product-line design problem, such as the one Belloni et al. (2008) used in their model. The model by Belloni et al. (2008) does not take into account such peer influence effects and assumes an a priori preference ordering among the different product profiles is available and therefore cannot be used to model this setting. Wang et al. (2009) also model the product-line design problem for the share-of-choice problem without requiring an a priori preference ordering among the product profiles for the consumers, however, their model does not require identification of the highest utility product. Determining the highest utility product is important in the setting where peer influences vary across products. Thus, we significantly expand upon the Wang et al. (2009) model by incorporating the peer influence effects and ensure the customer picks the product with the highest utility.[4]

To include the first-order network effects (a person is positively affected from a neighbor who uses the same product) we construct the following model. Let $M$ be the number of products in the product-line, then purchase among multiple products are represented by the additional decision variable type $y_{sq}$ which is 1 if person $s$ buys the $q$-th product in the product line. $x_{klq}$ is 1 if level $l$ is selected for attribute $k$ of the $q$-th product and is 0, otherwise. The objective of the problem maximizes the total number of buyers. Constraints (24) and (25) correspond to constraint (6) in the SOCSNE model and include only the first-order network effects. One level is selected for each attribute with constraint (26) and at most one product is allowed to be purchased by a customer with constraint (27), represented by the binary variable $y_s$ which is 1 if person $s$ buys a product and 0, otherwise. As mentioned above, we need to ensure a customer purchases the product that provides them the highest utility after taking social network effects into account. This is somewhat tricky to model and the constraint group (28) guarantees that if the utility from the $q$-th product is greater than the utility from the $r$-th product for a customer then the $q$-th product is preferred over the $r$-th product by that person. Here $B$ is a positive large integer number, though it is easily seen that it is bounded by the total number of levels and $c^s_{qr}$ is a binary variable required for the logical argument. All of the decision variables are binary variables. The share-of-choice problem with network effects for the product-line design problem (SOCNEPL) is given below.

$$\textbf{SOCNEPL:} \quad \text{Maximize} \quad \sum_{s=1}^{n} y_s, \tag{23}$$

$$\text{subject to} \quad \sum_{k=1}^{K}\sum_{l=1}^{L_k} u^s_{kl} x_{klq} + \Delta^1_s \sum_{j\in V} a_{js} y_{jq} = U_{sq} \qquad s=1,2,\ldots,n,\ q=1,2,\ldots,M, \tag{24}$$

$$h^H_s y_{sq} \le U_{sq} \qquad s=1,2,\ldots,n,\ q=1,2,\ldots,M, \tag{25}$$

$$\sum_{l=1}^{L_k} x_{klq} = 1 \qquad k=1,2,\ldots,K,\ q=1,2,\ldots,M, \tag{26}$$

$$\sum_{q=1}^{M} y_{sq} = y_s \qquad s=1,2,\ldots,n, \tag{27}$$

[4] We should note that the branch-and-price approach in Wang et al. (2009) relies on the procedure developed in Camm et al. (2006) to solve the pricing problem. As explained in Section 4 that approach cannot be applied in the setting with peer influence effects.

$$\left.\begin{aligned} U_{sq} - U_{sr} &\le B(1 - c^s_{qr}) \\ U_{sr} - U_{sq} &\le B c^s_{qr} \\ y_{sq} - y_{sr} &\le (1 - c^s_{qr}) \\ y_{sr} - y_{sq} &\le c^s_{qr} \\ c^s_{qr} &\quad \text{binary} \end{aligned}\right\} \begin{array}{l} s = 1, 2, \ldots, n, \\ \forall\, q, r \text{ pairs} \in \{1, 2, \ldots, M\}, \end{array} \tag{28}$$

$$y_s, y_{sq}, x_{klq} \text{ binary} \qquad \begin{array}{l} s = 1, 2, \ldots, n,\ l = 1, 2, \ldots, L_k, \\ k = 1, 2, \ldots, K, q = 1, 2, \ldots, M. \end{array} \tag{29}$$

The second-order network effects include buyers of other products in the same product-line. To differentiate the two dimensional network effects, we introduce a second-order effect, $\Delta_s^2$, which is of a smaller magnitude than the first order effect, $\Delta_s^1$. So the utility person $s$ gets from using product $q$ represented by $U_{sq}$ in constraint (24) would now be replaced with the following ($\Delta_s^2$ is subtracted from $\Delta_s^1$ to eliminate double counting of the effect from the same consumer).

$$\sum_{k=1}^{K} \sum_{l=1}^{L_k} u^s_{kl} x_{klq} + (\Delta_s^1 - \Delta_s^2) \sum_{j \in V} a_{js} y_{jq} + \Delta_s^2 \sum_{j \in V} a_{js} y_j = U_{sq}. \tag{30}$$

The SOCNEPL model is not computationally tractable as the LP relaxation is weak. We discuss how the GA can be modified for the product-line design problem. First, the representation of an individual in the population will need to include the different products in a line. For a more thorough search, the ordering of the products in the representation of an individual should be selected carefully. Once an initial population is selected, fitness is calculated using the new integer program, SOCNEPL. The crossover produces offspring for each product profile, so the single point crossover should be applied at every product. Similarly, in the mutation step a level is changed with a predetermined probability for each product. At the end of the mutation, when the population size is doubled, it is halved by selecting the product lines which have the total largest fitness. Clearly for the GA to be computationally tractable, the SOCNEPL model needs to be improved or an approximate heuristic procedure needs to be developed to compute the market share for a given product-line. This is the focus of our current research.

### 6.4. Application of the LCIP in Epidemiology and a Slight Generalization

Although the LCIP has been framed in a product diffusion setting, it can also be equivalently viewed in an epidemiological setting. Suppose that $e_{js}$ denotes the risk factors or influence of untreated node $j$ on node $s$ (e.g., if $\delta_{js}$ denotes the probability of node $s$ getting infected by untreated node $j$, then $e_{js} = -\log(1 - \delta_{js})$). Let $f_{js}$ denote the reduction of influence of node $j$ on node $s$ if node $j$ is treated so that its risk level is less than or equal to a threshold risk level

$r_j$. In other words, if node $j$ is treated so that its risk level is below $r_j$, its influence on node $s$ is $e_{js} - f_{js}$. We would like to ensure that the sum of all $e_{js} - f_{js}$'s for node $s$ minus the intervention or treatment strategy $z_s$ reduces the overall risk of node $s$ below the threshold risk level $r_s$. This may be equivalently cast in the marketing setting with $b_s = -r_s + \sum_{j \in V'} e_{js}$ and $d_{js} = f_{js}$, with a discrete set of intervention or treatment strategy choices at each node (with associated costs). The LCIP in the epidemiological setting is then the problem of finding a least cost treatment plan to ensure that a given population has its risk levels for a particular epidemic reduced to below a target threshold level for each member of the population.

This illustrates that the LCIP is an extremely useful model in a social network setting where the behavior of one's immediate neighbors influences ones own, and it is of interest to understand how "information" (loosely defined) spreads through a network. In particular, it is of interest to understand the power of different nodes in a network in terms of their relative "influence" in helping spread (or stop the spread) the "information" over a network.

In a setting where it is desired that a fraction of the population (as opposed to the entire population) be influenced and adopt the product the LCIP model is easily modified by replacing constraint (14) by $\sum_{s \in V'} y_{sT} = \alpha |V'|$; where $\alpha$ represents the desired fraction of individuals in the network that eventually adopt the product. When $\alpha = 1$ we obtain the previously defined LCIP problem.

## 7. Concluding Remarks

We proposed a novel model to include peer influence effects in product design within the framework of the share-of-choice (SOC) problem. Although the share-of-choice problem has been studied in the marketing literature, to our knowledge the peer influence effects have never been explicitly considered previously within the product design process. In this model, we attempt to develop a new product taking social network effects into account, before intervening during the marketing phase with targeted promotions. By taking into account peer influence effects, with our new model one is able to design products with far larger market shares than obtained by the original SOC model. While the effects of peer influence on consumer choice are well documented, previous analysis of conjoint data typically assumed that a consumer's attribute preferences and product choices are independent of choices of others. Narayan et al. (2011) take a significant first step in the development of conjoint estimation models that incorporate peer influence. The SOCSNE model introduced in this paper allows one to successfully use the results of such conjoint estimation models towards a logical next step—the design of a product with largest market share.

The new model we constructed remains a computationally challenging NP-Hard problem. However, after extensive computational studies we show that the GA used to solve this problem finds high quality solutions for the simulated data sets. Such evaluation is possible for small size data sets where the optimal is available for comparison. For the larger data sets, where CPLEX requires hours to solve the problem or is unable to do so, the GA is robust and preserves a running time around less than two minutes independent of the size of the problem. As one of the characteristics of the GA, it is flexible with extensions requiring only minor modifications to the algorithm.

Engineering the diffusion of a product is already used by businesses in the form of free samples or coupons. The modeling of this process (LCIP) by using the product profile obtained by the SOC-SNE model, and then intervening with tailored incentives for a group of customers to address the tradeoff between a larger market share and the costs of providing incentives is a second contribution of this paper. We model the LCIP using an integer program. Since this model is computationally very hard to solve, we develop a preprocessing procedure and a simple iterative strategy to solve it. As noted earlier the LCIP is applicable in a larger number of settings. For instance, while it is considered in the context of this paper within the framework of the product design process, in practice it may be used tactically and operationally during the marketing phase. Specifically, the population to which the product is marketed to is typically larger than the population used for the conjoint study. Further, social network structures are dynamic and may change over time. Thus, with a product (or product line) in place a marketer could benefit from the LCIP model to analyze the tradeoffs and incentives required to reach a desired fraction of the population. Although the LCIP model considers incentives in terms of customer utilities, one should note that following Miller et al. (2011) it is possible to infer dollar values from customer utilities.

There are several natural directions for future research. Broadly, they all encompass the expansion of product or product-line design problems to settings with social network or peer influence effects. They include (i) pricing of the levels of attributes for a product as part of the design process, (ii) consideration of alternate assumptions of product selection amongst consumers (i.e., instead of the highest choice pattern where each consumer deterministically self-selects the product from the line that provides her the highest surplus one could consider the multinomial choice logit model (Chen and Hausman 2000)), and (iii) extension of the problem to stochastic network settings (i.e., one in which the social network changes over time).

## References

Aral, S., D. Walker. 2011. Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science* Forthcoming.

Balakrishnan, P.V., V.S. Jacob. 1996. Genetic algorithms for product design. *Management Science* **42**(8) 1105–1117.

Bearden, W.O., M.J. Etzel. 1982. Reference group influence on product and brand purchase decisions. *Journal of Consumer Research* **9**(2) 183–194.

Bell, D.R., S. Song. 2007. Neighborhood effects and trial on the internet: Evidence from online grocery retailing. *Quantitative Marketing and Economics* **5**(4) 361–400.

Belloni, A., R. Freund, M. Selove, D. Simester. 2008. Optimizing product line designs: Efficient methods and comparisons. *Management Science* **54**(9) 1544–1552.

Bourne, F.S. 1957. Group influence in marketing and public relations. R. Likert, S.P. Hayes, eds., *Some Applications of Behavioral Research*. Basil, Switzerland: UNESCO, 208–224.

Burnkrant, R.E., A. Cousineau. 1975. Informational and normative social influence in buyer behavior. *Journal of Consumer Research* **2**(3) 206–215.

Camm, J.D., J.J. Cochran, D.J. Curry, S. Kannan. 2006. Conjoint optimization: An exact branch-and-bound algorithm for the share-of-choice problem. *Management Science* **52**(3) 435–447.

Chen, K.D., W.H. Hausman. 2000. Mathematical properties of the optimal product line selection problem using choice-based conjoint analysis. *Management Science* **46**(2) 327–332.

Childers, T.L., A.R. Rao. 1992. The influence of familial and peer-based reference groups on consumer decisions. *Journal of Consumer Research* **19**(2) 198–211.

Dobson, G., S. Kalish. 1993. Heuristics for pricing and positioning a product-line using conjoint and cost data. *Management Science* **39**(2) 160–175.

Goldberg, D.E. 1989. *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley, Reading, MA.

Goldenberg, J., B. Libai, E. Muller. 2001. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters* **12**(3) 211–223.

Granovetter, M. 1978. Threshold models of collective behavior. *American Journal of Sociology* **83**(6) 1420–1443.

Green, P.E., A.M. Krieger. 1989. Recent contributions to optimal product positioning and buyer segmentation. *European Journal of Operational Research* **41**(2) 127–141.

Green, P.E., V.R. Rao. 1971. Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research* **8**(3) 355–363.

Gupta, S., R. Kohli. 1990. Designing products and services for consumer welfare: Theoretical and empirical issues. *Marketing Science* **9**(3) 230–246.

Holland, J.H. 1975. *Adaptation in natural and artificial systems*. The University of Michigan Press, Ann Arbor, MI.

Iyengar, R., S. Han, S. Gupta. 2009. Do friends influence purchases in a social network? *Harvard Business School Marketing Unit Working Paper Num. 09-123* .

Iyengar, R., C. Van den Bulte, T.W. Valente. 2011. Opinion leadership and social contagion in new product diffusion. *Marketing Science* **30**(2) 195–212.

Katz, M.L., C. Shapiro. 1985. Network externalities, competition, and compatibility. *The American Economic Review* **75**(3) 424–440.

Kempe, D., J. Kleinberg, É. Tardos. 2003. Maximizing the spread of influence through a social network. *Proceedings of the ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 137–146.

Kohli, R., R. Krishnamurti. 1987. A heuristic approach to product design. *Management Science* **33**(12) 1523–1533.

Kohli, R., R. Krishnamurti. 1989. Optimal product design using conjoint analysis: Computational complexity and algorithms. *European Journal of Operational Research* **40**(2) 186–195.

Leenders, R.T., J. van Engelen, J. Kratzer. 2003. Virtuality, communication, and new product team creativity: A social network perspective. *Journal of Engineering and Technology Management* **20**(1–2) 69–92.

Manchanda, P., Y. Xie, N. Youn. 2008. The role of targeted communication and contagion in product adoption. *Marketing Science* **27**(6) 961–976.

Maniezzo, V., T. Stützle, S. Voß. 2009. *Matheuristics: Hybridizing Metaheuristics and Mathematical Programming*, vol. 10. Springer Verlag.

Michalewicz, Z. 1996. *Genetic algorithms + data structures = Evolution programs*. Artificial intelligence, Springer Verlag, Berlin.

Miller, K.M., R. Hofstetter, H. Krohmer, Z.J. Zhang. 2011. How should consumers' willingness to pay be measured? An empirical comparison of state-of-the-art approaches. *Journal of Marketing Research* **48**(1) 172–184.

Nair, S.K., L.S. Thakur, K.W. Wen. 1995. Near optimal solutions for product line design and selection: Beam search heuristics. *Management Science* **41**(5) 767–785.

Narayan, V., V.R. Rao, C. Saunders. 2011. How peer influence affects attribute preferences: A Bayesian updating mechanism. *Marketing Science* **30**(2) 368–384.

R Development Core Team. 2010. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org`. ISBN 3-900051-07-0.

Sawtooth Software. 2003. Advanced simulation module for product optimization v1.5. Tech. rep., Sawtooth Software, Inc., Sequim, WA.

Shi, L., S. Ólafsson, Q. Chen. 2001. An optimization framework for product design. *Management Science* **47**(12) 1681–1692.

Sosa, M.E., S.D. Eppinger, C.M. Rowles. 2004. The misalignment of product architecture and organizational structure in complex product development. *Management Science* **50**(12) 1674–1689.

Srinivasan, R., G.L. Lilien, A. Rangaswamy. 2004. First in, first out? The effects of network externalities on pioneer survival. *Journal of Marketing* **68**(1) 41–58.

Story, L., B. Stone. 2007. Facebook retreats on online tracking. `http://tinyurl.com/4358cev`. *The New York Times*.

Van den Bulte, C., S. Wuyts. 2007. *Social networks and marketing*. Marketing Science Institute, Cambridge, MA.

Wang, X.J., J.D. Camm, D.J. Curry. 2009. A branch-and-price approach to the share-of-choice product line design problem. *Management Science* **55**(10) 1718–1728.

Watts, D.J., S.H. Strogatz. 1998. Collective dynamics of "small-world" networks. *Nature* **393** 440–442.