Incentive-Compatible Advertising on a Social Network*

Kfir Eliaz and Ran Spiegler†

March 15, 2016

Abstract

A platform that operates a social network allows firms to post display ads to network members. Each member is interested in exactly one type of product. The network structure is correlated with the profile of members' privately known preferences over product types. The platform's policy consists of a display rule (which specifies the stationary probability with which each product is shown to each network member, as a function of the network structure) and an advertising fee (which the platform charges from firms as a function of their reported type). We provide conditions for the existence of an incentive-compatible policy that maximizes and fully extracts firms' surplus. This objective is easier to attain when the network is more informative of members' preferences, consumers are more attentive to advertising and their frequency of repeated purchases is higher, and advertisers are less informed of the network structure. We provide a more detailed characterization when the network is generated according to the "stochastic block model", thus linking our model to the "community detection" problem in Network Science.

*Financial support from ISF grant, no. 1153/13, is gratefully acknowledged. We thank Nikita Roketskyi and Omer Tamuz for helpful comments.

†Eliaz: School of Economics, Tel-Aviv University and Economics Dept., University of Michigan, kfire@post.tau.ac.il. Spiegler: School of Economics, Tel-Aviv University and Economics Dept., University College London, rani@post.tau.ac.il.

1 Introduction

The rise of modern online platforms has generated new opportunities for advertisers in terms of personalizing and targeting their marketing messages. When consumers access a platform, they leave a trail of information that may be correlated with their consumption tastes. This correlation enables advertisers to achieve better targeting, which in turn helps the platform increase its advertising revenues. In this paper, we study targeted advertising when the platform operates a *social network*. Consumers access the platform in order to cultivate relationships with other network members - i.e., the platform's primary function is not commercial but social. However, consumers' social activity generates valuable information for advertisers. In particular, the structure of the social network, and the exact location of an individual consumer in the network are informative of his preferences. For instance, if individuals exhibit *homophily* - i.e., they associate with other individuals with similar interests and tastes (see Kandell (1978)) - then a large cluster in the network indicates that its members are likely to have similar tastes.

Our objective is to develop a modeling framework that addresses a number of key questions: How can the platform use the information inherent in the network's structure to allocate display-ad slots to advertisers? What are the incentive constraints that the platform faces? When can the platform overcome these incentive constraints and fully extract advertisers' surplus? How is implementability of the platform's objective affected by various characteristics of its environment: the distribution of consumer tastes, the way these tastes are correlated with the formation of social links, consumers' attentiveness to advertising, the frequency of repeat purchases and the extent to which advertisers are informed of the network structure?

In our model, there are two product types, x and y. Each type is offered by a large number of firms (also referred to as advertisers). There is a collection of n consumers. The type of each consumer is defined by the one product type he is interested in. Consumers' privately known types are drawn *i.i.d.* Consumers are linked by a social network, the structure of which is a stochastic function of the profile of consumers' types. The social network is operated by a platform that enables firms to post personalized display ads. Firms know the process that generates the network. As a benchmark, we assume that they are uninformed of the realized network - we later relax this assumption.

We consider a stationary environment in which each consumer type is in one of two states - a "demand" state in which he is potentially interested in consuming his product type, and a "satiation" state in which the consumer is not interested in consumption. A consumer in the demand state switches to his satiation state as soon as he consumes his product type. He switches back to the demand state with per-period probability ε.

At every time period, each consumer is exposed to a display ad that belongs to one of the two product categories, x or y. While consumers have continual access to the platform, their attention to advertising on the platform is limited: each consumer notices his personalized display ad with probability θ at each period. When the consumer notices an ad, he realizes whether it describes a product he likes. If it does, then a consumer in his demand state transacts with the advertiser "offline" - i.e., outside the platform - whereas a consumer in his satiation state does not transact. The consumer's presence at the platform continues uninterrupted even after he transacts because his primary motivation for accessing it is the social network it operates. We abstract from product prices, and assume that each firm aims to maximize its expected per-period number of transactions.

The platform's knowledge of the social network enables it to refine its beliefs regarding each consumer's type. Therefore, although consumers are ex-ante identical, the platform is able to tailor its ad display policy to individual consumers, according to the information revealed by the realized network. The platform employs a *stationary (personalized) display rule*, such that the product category that is displayed to a consumer is drawn independently at every period, according to a distribution that is allowed to vary across consumers. Stationary display rules are justified by the assumption that consumers have continual access (from "period minus infinity") to the social network, and that the platform cannot monitor whether consumers transact with firms or pay attention to their ads. The platform's policy consists of its chosen display rule and an advertising fee, which is a per-period lump sum that is allowed to depend on the firm's type.

If the platform could identify firms' types, it would choose a policy that maximizes and fully extracts firms' surplus (i.e., their per-period number of transactions). Because of the uncertainty regarding consumer types and the restriction to stationary display rules, the platform would find it optimal to use *interior* rules - i.e., it would display *both products* with positive probability to some consumers and network realizations - as long as ε/θ is not too high. At the other extreme, when ε/θ is high enough, the optimal display rule is deterministic: each consumer will be exposed to exactly one type of ad.

The key assumption in the paper, however, is that the platform is unable to directly verify advertisers' types. Therefore, the display-ad slots it allocates to each firm (according to its display rule) and the advertising fee it charges are based on the firm's *self-reported* type. To give a sense of what such self-reporting corresponds to in real life, imagine that x and y represent different categories of holiday packages: resort vs. active travel. In this context, reporting one's type means self-selecting one of the two categories for the purpose of getting advertising slots. Thus, advertisers communicate to the platform which category of ad slots to apply to them. The probability with which an advertiser will be displayed to members of the network will depend on its selected category, and our question is whether advertisers will have an incentive to select the appropriate category. The platform's policy is incentive-compatible if every advertiser wants to report truthfully, given that all other advertisers do so. Our main problem is to characterize the situations in which the platform's ideal policy (a.k.a its first-best) is incentive-compatible.

It turns out that the possible optimality of interior display rules creates an incentive issue that would not arise under deterministic rules. Our basic result is a necessary and sufficient condition for the implementability of the first-best when ε/θ is low (such that the optimal display rule is interior for *all* consumers and network realizations). Moreover, when ε/θ is sufficiently low, the assumption that advertisers do not know the realized network is important: If they were *fully* informed of the network structure, the platform would be unable to implement the first-best.

The necessary and sufficient condition is a simple inequality that incorporates two quantities: (i) the extent to which the network is informative of consumers' types - measured by the *Bhattacharyya Coefficient* of similarity between the distributions over realized networks conditional on the two possible consumer types; and (ii) the ratio between the ex-ante probabilities that consumers like x and y. As the network becomes more informative, and as the distribution over consumer tastes becomes more symmetric, it gets easier to meet the condition. However, quantities (i) and (ii) are not independent: the informativeness measure is sensitive to the distribution over consumer types. This interdependence leads to a few surprising conclusions, as we will see below.

The necessary and sufficient condition enables us to address some of the comparative-statics questions raised in the opening paragraphs. An increase in consumers' attentiveness, or a decrease in the frequency of repeat purchases, makes it harder to satisfy the necessary condition and therefore *magnifies* the incentive constraints that limits the platform's ability to implement its first-best. The reason is that a lower ratio ε/θ pushes optimal display probabilities away from the extremes, thus exacerbating the incentive issues that arise from interior display rules.

While our basic result is straightforward to derive, working out its ramifications is non-trivial, because of the intricacies of calculating the Bhattacharyya Coefficient induced by the joint distribution over consumer-type profiles and social networks. We illustrate the basic result - and specifically, the interplay between the above factors (i) and (ii) - with a familiar model of randomly generated networks, known as the "*stochastic block model*". This model has been extensively studied in the Computer Science literature (e.g., see Mossel, Neeman and Sly (2012) and Abbe and Sandon (2015)). It is parameterized by the size of the network n and a "connectivity matrix" - i.e., a symmetric 2×2 matrix that specifies the probability of a link between any pair of consumers as a function of their types. The stochastic block model subsumes two natural models of network formation: the "homophily" case where identical types are more likely to form a link, and the "extroversion/introversion" case where each consumer type has a different propensity to initiate a link, and the probability of a link between two consumer types is a product of their propensities.

Applying our main result to the stochastic block model, we show that when ε/θ is small enough, the first-best is not implementable if the consumer type distribution is either too biased in favor of one type or too close to being uniform. We also use the stochastic block model to address the following question. Suppose that the platform cannot prevent advertisers from learning the subgraph induced by the social network over a random subset of m nodes; how large can m be without destroying the first-best's implementability? We present a sufficient condition that quantifies the upper bound on m in terms of the consumer type distribution and the Bhattacharyya Coefficient induced by the stochastic block model with $n - m$ nodes.

A natural question for the stochastic block model is whether a larger network makes it easier for the platform to implement its first-best. We first present a simple result: for a fixed connectivity matrix, the first-best is implementable if n is sufficiently large. This is because the network becomes arbitrarily informative in the $n \to \infty$ limit. The computer Science literature has addressed a substantially more difficult question. Suppose that we raise n and lower the connectivity matrix at the same time, such that the expected degree of an individual node grows only *logarithmically* in n; does the network become arbitrarily informative in the $n \to \infty$ limit? Recently, Abbe and Sandon (2015) derived a characterization of the connectivity matrices for which the answer is affirmative. We apply their result to obtain a sufficient condition for the implementability of the first-best, and illustrate it for the homophily and extroversion/introversion specifications.

Related literature

This paper belongs to a research agenda that explores novel incentive issues that arise in modern online platforms. Our earlier exercise in this vein, Eliaz and Spiegler (2015), studied an environment in which consumers submit noisy queries regarding their preferences. A "search platform" responds to each query by providing consumers with a "search pool" - i.e., a collection of products that they can browse via some search process. The platform's problem is to design a decentralized mechanism for efficiently allocating firms into these search pools, and for extracting the firms' surplus. Some of the ingredients of our earlier exercise - notably the relevance of the above factors (i) and (ii) for the implementation problem - reappear in the present model, albeit for different reasons and in a somewhat different form. However, new substantive and technical questions arise because of the specific context of social networks.

There has been a growing interest in targeted advertising in the I.O. literature. One strand of this literature analyzes competition between advertising firms that choose advertising intensity, taking into account the cost of advertising and the probability that their advertising messages will reach the targeted consumers. Notable papers in this literature include Iyer, Soberman and Villas-Boas (2005), Athey and Gans (2010), Bergemann and Bonatti (2011), Zubcsek and Sarvary (2011) and Johnson (2013). A second strand of this literature studies how to optimally propagate information about a new product by targeting specific individuals in a network. Recent papers in this strand include Galeotti and Goyal (2012) and Campbell (2015) (see Bloch (2015) for a survey).

2 A Model

Let $N = \{1, ..., n\}$ be a set of consumers (nodes), and let $T = \{x, y\}$ be a set of product types. Each product is offered by a number m of advertisers who can costlessly supply any amount of this product. Each consumer is interested in exactly one product type. We say that a consumer (advertiser) is of type x (y) if he (it) demands (supplies) product x (y). We use t_i to denote the type of consumer $i \in N$. If a consumer acquires a product from an advertiser, the advertiser earns a fixed payoff of 1 (we abstract from product pricing).[1]

Consumers and advertisers are brought together by a platform that operates a social network - i.e., it enables consumers to form social links with each other. Whether a pair

[1] It is easy to adapt our analysis to the case of profit margins that vary across product types. We assume symmetry across product types purely for notational simplicity.

of consumers is linked depends stochastically on their types. The social links between consumers define a network w, which is a non-directed graph over the set of nodes N. The set of all possible networks is W. From now on, we will refer to elements in N as consumers and nodes interchangeably.

Let μ be the joint distribution over the profile of consumer types and the social network. We use $\mu_i(t, w)$ to denote the probability that a given consumer i is of type t and the network is w. Let $\mu_i(x) = \sum_{w \in W} \mu_i(x, w)$ be the ex-ante probability that $t_i = x$. Given some network w, the probability that $t_i = x$ is denoted $\mu_i(x \mid w)$. Likewise, $\mu(w \mid t_i)$ is the distribution over networks conditional on consumer i's type. We assume that μ is *symmetric* in the following sense. First, all consumers are ex-ante identical - i.e., $\mu_i(x) = \mu(x)$ for every $i \in N$. Second, suppose nodes i and j are indistinguishable in the network w - i.e., any node $h \neq i, j$ is linked to i if and only if it is linked to j. Then, $\mu_i(x \mid w) = \mu_j(x \mid w)$. Finally, denote $\mu(x) = \pi \geq \frac{1}{2}$.

We consider a stationary environment in which a consumer of type t is in one of two states: state D_t, a "demand state" in which the consumer purchases product type t whenever he notices an ad for it (we describe below the process by which ads are displayed to consumers) and a state S_t, a "satiation state" in which the consumer is not interested in consuming. A consumer who is in state D_t remains in that state until he consumes product type t, in which case he switches to state S_t. A consumer who is in state S_t returns to state D_t with independent per-period probability ε. Thus, the parameter ε captures the frequency with which consumers are active in the product market.

We assume that the platform observes the network, but does not directly observe the types of consumers and advertisers. Each consumer and each advertiser only knows its own type. In particular, advertisers do not observe the realized social network. We will relax the latter assumption in Section 5.

An example: A three-node network with perfect homophily
The following specification will serve as a running example in the paper. Let $n = 3$ and assume that nodes i and j are linked in w if and only if $t_i = t_j$. Then, the network is pinned down by the profile of consumer types. In particular, the only networks that are realized with positive probability are the fully connected graph and a graph in which exactly two nodes are connected. We can use this observation to calculate $\mu_i(w \mid t_i)$. For example, the probability that the network is fully connected conditional on $t_1 = x$ is π^2, while the probability of this network conditional on $t_1 = y$ is $(1 - \pi)^2$; the probability of the network in which only 1 and 2 are linked conditional on $t_1 = x$ is $\pi(1 - \pi)$; and so forth. $\square$

How does the platform match advertisers and consumers? At every time period and for each consumer i, the platform selects an advertiser - according to a process we will specify below - and displays it to the consumer in the form of an advertising banner. The consumer notices the ad with probability $\theta > 0$ (independently across periods). The parameter θ captures the consumer's *attention* to advertising on the social network. If the consumer fails to notice the ad, it expires at the end of the period and a new one is displayed in the next period. When the consumer notices the ad, he instantaneously recognizes the advertised product. If it does not match his type, the consumer takes no action and stays in the market where his exposure to ads continues. If the advertised product matches the consumer's type, he acquires it if and only if he is in his demand state. After the consumer buys the product, he switches to his satiation state but remains in the platform, as its primary function is the social network it operates.

The platform's objective is to maximize expected profits by assigning advertisers to nodes and charging advertisers a per-period lump-sum access fee. The advertisers' objective is to maximize their per-period expected profits. We assume that the platform controls which advertiser is displayed to any consumer at any time period. However, the platform cannot monitor whether the consumer is attentive to the ad, or whether a transaction is made. This fits situations in which all interactions between consumers and advertisers take place "offline" (i.e., outside the platform), and there is no notion of "clicking" on display ads.

3 Basic Results

In this section we characterize the policy that would maximize the platform's advertising revenues if it could observe advertisers' types, and then derive conditions for the incentive-compatibility of the optimal policy.

3.1 First-Best

The "first-best" outcome for the platform consists of an assignment of advertisers to network nodes that maximizes the advertisers' surplus, and a fee schedule that fully extracts this surplus. We assume that the platform precommits to personalized, stationary display rules. Stationarity is justified by the stationarity of the process that governs consumers' switching between demand and satiation states, and by the platform's inability to monitor consumers' commercial activity.

Formally, $q_i(t \mid w)$ is the probability that at any time period, the platform displays an advertiser of type $t \in \{x, y\}$ to consumer i, conditional on the realized network being w. Of course, $q_i(y \mid w) \equiv 1 - q_i(x \mid w)$. Conditional on displaying an advertiser of type t, each of these advertisers is drawn with equal probability. Hence, the probability that a particular advertiser of this type is displayed is $q_i(t \mid w)/m$. We refer to $q(w) = (q_i(x \mid w))_{i \in N}$ as the platform's *display rule* for w. Let F_t be the per-period fee the platform charges from advertisers of type t . Denote $q = (q(w))_{w \in W}$, $F = (F_x, F_y)$. The pair (q, F) constitutes the platform's *policy.*

Our objective is to characterize the policy that maximizes the platform's advertising revenues. For this purpose, let us first derive the collection of display rules that maximizes advertisers' surplus. The gross expected per-period payoff (without taking into account any payment to the platform) for an advertiser of type t is calculated as follows. For each consumer $i \in N$, we compute the invariant probability that the consumer is in state D_t, and we multiply it by the probability that the consumer transacts conditional on being in this state. Then, we sum over all consumers. To carry out this calculation, let us first derive the invariant distribution over states of each consumer type.

Given a network w and a display rule $q(w)$, the transition probabilities between the states of consumer i of type t are given by the following matrix:

$$\begin{array}{ccc} & D_t & S_t \\ D_t & 1 - \theta q_i(t \mid w) & \theta q_i(t \mid w) \\ S_t & \varepsilon & 1 - \varepsilon \end{array} \tag{1}$$

Hence, given w and $q(w)$, the joint invariant probability that consumer i is in state D_t is

$$\rho_t(i, w) \equiv \frac{\lambda \mu_i(t \mid w)}{\lambda + q_i(t \mid w)}$$

where

$$\lambda = \frac{\varepsilon}{\theta}$$

This parameter will play a key role in our analysis. It follows that the expected number of transactions per period with advertisers of type t is

$$\sum_{i \in N} \sum_{w \in W} \mu(w) \rho_t(i, w) \theta q_i(t \mid w) \tag{2}$$

Let $q^*(w)$ be the display rule that maximizes the sum

$$\sum_{i \in N} [\rho_x(i,w)\theta q_i(x \mid w) + \rho_y(i,w)\theta q_i(y \mid w)] \tag{3}$$

We refer to $q^*(w)$ as the "first-best" display rule for the network w. The "first-best" fee that the platform charges advertisers of type t, denoted F_t^*, is defined to be the amount that extracts the maximal expected surplus of these advertisers. I.e., F_t^* is equal to (2) evaluated at $q^*(w)$.

Note that for fixed μ, if consumers are sufficiently inattentive (i.e., θ is sufficiently close to zero such that λ is sufficiently high), the first-best display rule is generically a corner solution: $q_i^*(t \mid w) = 1$ if $\mu_i(t \mid w) > \frac{1}{2}$. When $q_i^*(w)$ is interior, first-order conditions imply

$$\frac{\lambda + q_i(x \mid w)}{\lambda + q_i(y \mid w)} = \sqrt{\frac{\mu_i(x \mid w)}{\mu_i(y \mid w)}} \tag{4}$$

This allows us to solve explicitly for $q_i^*(x \mid w)$:

$$q_i^*(x \mid w) = \frac{(1+\lambda)\sqrt{\mu_i(x \mid w)} - \lambda\sqrt{\mu_i(y \mid w)}}{\sqrt{\mu_i(y \mid w)} + \sqrt{\mu_i(x \mid w)}} \tag{5}$$

In particular,

$$q_i^*(x \mid w) \to \frac{\sqrt{\mu_i(x \mid w)}}{\sqrt{\mu_i(x \mid w)} + \sqrt{\mu_i(y \mid w)}} \tag{6}$$

when λ tends to zero.

Henceforth, we assume that the primitives μ and λ are such that

$$\frac{\lambda}{\lambda+1} < \sqrt{\frac{\mu_i(x \mid w)}{\mu_i(y \mid w)}} < \frac{\lambda+1}{\lambda} \tag{7}$$

for every w and every i, such that $q^*(w)$ is an interior solution for every w.

3.2 Incentive Compatibility

Let us now assume that the platform is unable to directly verify advertisers' types, and instead relies on their self-reports. A policy (q, F) is incentive-compatible (IC) if no firm has an incentive to misreport its type, given that every other firm reports truthfully. When a single advertiser of type t misreports its type, it changes its probability of display from $q_i(t \mid w)/m$ to $q_i(t' \mid m)/(m+1)$. Hence, the transition probability from D_t

to S_t changes to $\theta[q_i(t \mid w)+q_i(t' \mid m)/(m+1)]$, since this consumer will make a purchase if he is attentive and the displayed ad was either from one of the truthful x advertisers or from the single deviant advertiser. Consequently, the invariant probability that consumer i is in state D_t is

$$\rho_t'(i,w) = \frac{\lambda \mu_i(t \mid w)}{\lambda + q_i(t \mid w) + q_i(t' \mid w) \cdot \frac{1}{m+1}}$$

Note that $\rho_t' \to \rho_t$ as $m \to \infty$.

It follows that an x advertiser weakly prefers not to report that its type is y if and only if

$$\sum_{i\in N}\sum_{w\in W} \mu(w)\rho_x'(i,w)\theta\frac{q_i(x \mid w)}{m} - F_x \geq \sum_{i\in N}\sum_{w\in W} \mu(w)\rho_x'(i,w)\theta\frac{q_i(y \mid w)}{m+1} - F_y$$

We refer to this inequality as the $IC(x,y)$ constraint. The IC constraint of a y advertiser, or $IC(y,x)$, is similarly defined.

We wish to derive conditions under which the first-best policy (q^*, F^*) is IC in the $m \to \infty$ limit. When it is, we say that the first-best is implementable. Because F^* fully extracts advertisers' surplus, the L.H.S of $IC(x,y)$ and $IC(y,x)$ is zero. In the $m \to \infty$ limit, the inequalities thus reduce to

$$\begin{aligned} &\sum_{i\in N}\sum_{w\in W} \mu(w)q_i(y \mid w)(\rho_x(i,w) - \rho_y(i,w)) \leq 0 \\ &\sum_{i\in N}\sum_{w\in W} \mu(w)q_i(x \mid w)(\rho_y(i,w) - \rho_x(i,w)) \leq 0 \end{aligned} \tag{8}$$

Plugging the solution for q^* from the previous sub-section, we can obtain a necessary and sufficient condition for implementability of the first-best. However, in order to present this condition in an interpretable, transparent form, we need to introduce a new concept.

The Bhattacharyya Coefficient

Suppose that we knew the type of a particular consumer i. Then, we could update our beliefs regarding the overall structure of the social network. The conditional distributions $(\mu_i(w \mid t_i)_{w\in W}$, $t_i = x, y$, describe these updated beliefs. The following measure of similarity between these two conditional distributions turns out to play a key role in the condition for implementability of the first-best. Define

$$S_i \equiv \sum_{w \in W} \sqrt{\mu_i(w \mid x)\mu_i(w \mid y)}$$

In the Statistics and Machine Learning literatures, S_i is known as the *Bhattacharyya Coefficient* that characterizes the distributions $\mu_i(\cdot \mid x)$ and $\mu_i(\cdot \mid y)$.[2] From a geometric point of view, this is an appropriate similarity measure because S_i is the direction cosine between two unit vectors in $\mathbb{R}^{|W|}$, $(\sqrt{\mu_i(w \mid x)})_{w\in W}$ and $(\sqrt{\mu_i(w \mid y)})_{w\in W}$. The value of S_i increases as the angle between these two vectors shrinks; $S_i = 1$ if the two vectors coincide; and $S_i = 0$ if they are orthogonal.

More importantly, $S_i(x, y)$ is an appropriate similarity measure given our interpretation of $\mu_i(\cdot \mid x)$ and $\mu_i(\cdot \mid y)$, according to which the realized network w serves as a signal that indicates the types of individual consumers. Indeed, the stochastic matrix $(\mu_i(\cdot \mid t))_{t\in\{x,y\}}$ can be viewed as an information system in Blackwell's sense. The following result (which is stated and proved in Eliaz and Spiegler (2015)) establishes a link between Blackwell informativeness and the Bhattacharyya Coefficient.

Remark 1 *The Bhattacharyya Coefficient S_i decreases with the Blackwell informativeness of $(\mu_i(\cdot \mid t))_{t\in\{x,y\}}$.*

We will make extensive use of this observation in Sections 4 and 5. Finally, note that by the symmetry assumptions we imposed on μ, S_i is the same for all nodes, hence in what follows we will suppress the subscript i.

The following remark states two additional useful properties of the Bhattacharyya Coefficient.

Remark 2 *Let $w = (g, h)$. For each $k = g, h$ and collection of conditional distributions $(p(k \mid t))_{k,t}$, define*

$$S_p(k) = \sum_{w_k} \sqrt{p(k \mid x)p(k \mid y)}$$

(i) If $\mu(w \mid t) = \mu(g \mid t)\mu(h \mid t)$, then $S = S_\mu(g)S_\mu(h)$.
(ii) If $\mu(w \mid t) = \alpha(g)\beta^g(h \mid t)$, then $S = \sum_g \alpha(g)S_{\beta^g}(h)$.

[2] See Basu, Shioya and Park (2011) and Theodoris and Koutroumbas (2008). A related concept is the *Hellinger distance* between distributions, given by $H^2(x, y) = 1 - \sqrt{S(x, y)}$.

Part (i) says that the Bhattacharyya Coefficient induced by a collection of signals that are independent conditional on the consumer's type is the *product* of the Bhattacharyya Coefficients of the individual signals. Part (ii) says that when one signal is an independently distributed variable that merely determines the distribution of the other signal conditional on the consumer's type, the Bhattacharyya Coefficient is a weighted average of the coefficients induced by the latter signal. The properties follow immediately from the coefficient's definition, and therefore the proof is omitted.

To illustrate the Bhattacharyya Coefficient in our context, let us revisit the three-node example of Section 2. Let w_{ijl} denote the fully connected network, and let w_{ij} denote the network in which only nodes i and j are linked. Then,

$$S = \sqrt{\mu_i(w_{ijl} \mid x)\mu_{i1}(w_{ijl} \mid y)} + \sqrt{\mu_i(w_{jl} \mid x)\mu_i(w_{jl} \mid y)} + 2\sqrt{\mu_i(w_{ij} \mid x)\mu_i(w_{ij} \mid y)}$$

In Section 2, we derived the values of the conditional probabilities that feature in this expression. Plugging these values, we obtain $S = 4\pi(1 - \pi)$.

Necessary and sufficient conditions for first-best implementability

The next result employs the Bhattacharyya Coefficient to derive a simple statement of necessary and sufficient conditions for implementability of the first-best.

Proposition 1 *The first-best policy (q^*, F^*) is implementable if and only if*

$$S \leq (\frac{1+\lambda}{1+2\lambda})\sqrt{\frac{1-\pi}{\pi}} + (\frac{\lambda}{1+2\lambda})\sqrt{\frac{\pi}{1-\pi}} \tag{9}$$

Thus, implementability of the first-best depends on two factors: the two products' "popularity ratio" $\pi/(1-\pi)$, and the extent to which the social network is *informative* of consumer types, captured by the Bhattacharyya Coefficient induced by μ. To get an intuition for the result, consider the $\varepsilon \to 0$ limit, where repeat purchases by the same consumer are rare, and condition (9) simplifies into

$$S\sqrt{\frac{\pi}{1-\pi}} \leq 1 \tag{10}$$

In this parameter regime, the optimal display probability $q_i(t \mid w)$ is proportional to the *square root* of $\mu_i(t \mid w)$. By comparison, the first-best fee paid by a firm that submits the report t is proportional to $\mu(t)$. Thus, although a product with high $\mu_i(t \mid w)$

gets an advantage in terms of display probability, the square root factor *softens* this advantage.

The optimal policy's differential treatment of display probabilities and fees is a force that favors the less popular product y, thus creating an incentive for x firms to misreport. When $\pi/(1-\pi)$ gets larger (holding S fixed), the gap between the fees paid by the two types widens, and this exacerbates the misreporting incentive. As the network becomes more informative, the values of $\mu_i(t \mid w)$ get closer to zero or one, such that the "square root" effect vanishes, and this mitigates the misreporting incentive. Finally, recall that the platform conditions the display probabilities on w, whereas advertisers are uninformed of w at the time they submit their reports. When the network is highly informative, a firm that chooses to misreport knows it will be displayed with high (low) probability to consumers with low (high) probability of transacting with it, and this is another force that mitigates the misreporting incentive.

The above discussion may give an impression that the two factors, captured by S and $\pi/(1-\pi)$, are independent. This is not the case, because changes in the consumer type distribution generally lead to changes in the informativeness of the network. For example, recall that in our simple three-node network with perfect homophily, S is a strictly decreasing function of π (in the presumed $\pi \geq \frac{1}{2}$ range). In Section 4 we show that implementability of the first-best is not monotone in π in the sense that when π is close to 1 and also when π is close to $\frac{1}{2}$ the first-best is not implementable when the frequency of recurring purchases ε is small enough.

The probability ε of exiting the satiation state and the "attention parameter" θ contribute to the coefficient $\lambda/(1+2\lambda)$ that features in condition (9). Note that $\lambda/(1+2\lambda) \in (0, \frac{1}{2})$ and that it increases with λ, implying that it *increases* with ε and *decreases* with θ. The following result summarizes the comparative statics of the necessary condition w.r.t λ.

Proposition 2 *If the first-best is not implementable for a given λ, then it is not implementable under $\lambda' < \lambda$.*

This observation follows from noting that $\pi \geq \frac{1}{2}$. Thus, as consumers become *more attentive* to ads, or as the frequency of repeat purchases *declines*, the necessary and sufficient condition for implementing the first-best becomes *harder* to meet.

The results in this section only rely on the view of $\mu_i(\cdot \mid t)$ as a Blackwell information system. The fact that w is a social network plays no role, and in principle the same analysis would hold for an arbitrary signal w. Thus, one could argue that our insistence on the network interpretation is superfluous. However, note that there are few other concrete examples for an aggregate signal that provides information about the preferences of all consumers at the same time. But more importantly, in the next section we will pour more content into the network interpretation of w and use it to draw powerful implications from the basic results given in this section.

4 The Stochastic Block Model

Within the literature on random networks in various disciplines, a popular class of specifications is the *stochastic block model* (SBM). An SBM is characterized by a triplet (n, σ, P), where n is the number of nodes, σ is a probability vector over k types and P is a $k \times k$ symmetric matrix, where the entry P_{ij} gives the independent probability that a node of type i forms a link with a node of type j. In the case of two types ($k = 2$), the type distribution σ is represented by $\pi \in (0, 1)$ (the ex-ante probability that a node is of type x); and the connectivity matrix P is characterized by three parameters: p_x, the probability that two x types connect, p_y, the probability that two y types connect, and p_{xy} the probability that two different types connect. The components σ and P generate a joint distribution μ over consumer-type profiles and social networks that satisfies the symmetry properties we assumed in Section 2.

One of the central problems that is studied using SBMs is that of *community detection* (see Mossel, Neeman and Sly (2012) and Abbe and Sandon (2015), and the references therein). The objective is to identify with high probability the types of nodes in a given network, under the assumption that the network was generated by a known SBM. A growing literature in Computer Science and Machine Learning looks for conditions on the SBM parameters that are necessary and sufficient for identifying node types (and for implementing the identification with computationally efficient algorithms). These conditions capture the extent to which the network is informative about node types. Because this is also a crucial consideration for our problem of designing incentive-compatible advertising policies, the community-detection literature allows us to obtain simple sufficient conditions for implementability of the first-best, when the network formation process obeys an SBM.

When analyzing social networks, a natural question that arises is what induces two agents to form a link. One popular theory, known as *homophily,* is that agents with similar characteristics are more likely to connect. The connectivity matrix in this case can be captured by two parameters: $p_x = p_y = \alpha$, and $p_{xy} = \beta < \alpha$. An alternative theory is that some agents have a greater propensity to form social links than others.

We refer to this theory as *extroversion/introversion.* This case can also be represented with two parameters $\alpha > \beta$, such that $p_x = \alpha^2$, $p_y = \beta^2$ and $p_{xy} = \alpha\beta$.

Our first result in this section addresses the role of the consumer type distribution given by π. It turns out that when the popularity gap is too large or too small, the first-best is not implementable when λ is small.

Proposition 3 *Fix $n \geq 2$ and a generic P. (i) There exists $\pi^* \in (\frac{1}{2}, 1)$ with the property that for every $\pi > \pi^*$ there exists $\lambda^*(\pi)$ such that for every $\lambda < \lambda^*(\pi)$, the first-best is **not** implementable. (ii) There exists $\pi^{**} \in (\frac{1}{2}, 1)$ with the property that for every $\pi \in (\frac{1}{2}, \pi^{**})$ there exists $\lambda^{**}(\pi)$ such that for every $\lambda < \lambda^{**}(\pi)$, the first-best is **not** implementable.*

Proof. Our method of proof is to obtain two different lower bounds on S, and use these bounds to derive π^* and π^{**}.

(i) Fix a node i. Suppose that the platform were informed of the realized network w, as well as of t_j for all $j \neq i$. This would clearly be a (weakly) more informative signal of t_i than learning w only. Moreover, conditionally on learning $(t_j)_{j\neq i}$, the link status between any $j, h \neq i$ has no informational content regarding t_i. Therefore, in order to calculate a lower bound on S, we can consider a signal that consists of $(t_j)_{j\neq i}$ and the link status between i and every other j.

Let us calculate the Bhattacharyya Coefficient of the signal that consists of learning t_j and whether nodes i and j are linked:

$$\begin{aligned} & \sqrt{\pi p_x \cdot \pi p_{xy}} + \sqrt{\pi(1-p_x)\cdot\pi(1-p_{xy})} \\ & + \sqrt{(1-\pi)p_{xy}\cdot(1-\pi)p_y} + \sqrt{(1-\pi)(1-p_{xy})\cdot(1-\pi)(1-p_y)} \\ = \;& \pi\left(\sqrt{p_x p_{xy}} + \sqrt{(1-p_x)(1-p_{xy})}\right) + (1-\pi)\left(\sqrt{p_y p_{xy}} + \sqrt{(1-p_y)(1-p_{xy})}\right) \end{aligned}$$

Because signals that correspond to different nodes $j \neq i$ are independent conditional on t_i, the Bhattacharyya Coefficient of the signal that consists of $(t_j)_{j\neq i}$ and the link status between i and every other j is

$$\left[\pi\left(\sqrt{p_x p_{xy}} + \sqrt{(1-p_x)(1-p_{xy})}\right) + (1-\pi)\left(\sqrt{p_y p_{xy}} + \sqrt{(1-p_y)(1-p_{xy})}\right)\right]^{n-1}$$

Recall that by construction, this expression is weakly below S. Without loss of generality, let

$$\sqrt{p_x p_{xy}} + \sqrt{(1-p_x)(1-p_{xy})} \leq \sqrt{p_y p_{xy}} + \sqrt{(1-p_y)(1-p_{xy})}$$

Then, S is weakly above

$$\delta \equiv \left(\sqrt{p_x p_{xy}} + \sqrt{(1-p_x)(1-p_{xy})} \right)^{n-1}$$

For generic P (in particular, when all matrix entries get values in $(0,1)$), this term is strictly positive.

Let π^* satisfy $\sqrt{(1-\pi^*)/\pi^*} = \delta^2$, where $\delta < \frac{1}{2}$. For any $\pi > \pi^*$ let $\sqrt{(1-\pi)/\pi} = \hat{\delta}^2$ where $\hat{\delta} < \delta$. Define $\lambda^*(\pi)$ by $\lambda^*/(1+2\lambda^*) = \hat{\delta}^4$. Hence, for $\pi > \pi^*$ and $\lambda < \lambda^*(\pi)$ we have

$$(1 - \frac{\lambda}{1+2\lambda})(\frac{1-\pi}{\pi}) + (\frac{\lambda}{1+2\lambda})(\frac{\pi}{1-\pi}) < (1-\hat{\delta}^4)\hat{\delta}^4 + \hat{\delta}^4 \cdot \frac{1}{\hat{\delta}^2} < 2\hat{\delta}^2 < \delta$$

Therefore, there exists $\pi^* \in (\frac{1}{2}, 1)$ with the property that for every $\pi > \pi^*$ there exists $\lambda^*(\pi)$ such that for π and $\lambda \leq \lambda^*(\pi)$ we have

$$S \geq \left(\sqrt{p_x p_{xy}} + \sqrt{(1-p_x)(1-p_{xy})} \right)^{n-1} > (1 - \frac{\lambda}{1+2\lambda})(\frac{1-\pi}{\pi}) + (\frac{\lambda}{1+2\lambda})(\frac{\pi}{1-\pi})$$

which means that the necessary condition for implementability of the first-best is violated.

(ii) Let us now obtain a different lower bound on S. Once again, we use the fact that S decreases with the informativeness of the signal given by the network. For fixed n and π, this informativeness is maximal under perfect homophily - i.e., when $p_x = p_y = 1$ and $p_{xy} = 0$ - because any other connectivity matrix P represents a Blackwell garbling of it.

Assume perfect homophily, and consider an arbitrary node. Conditional on this node's type, if we learn whether it is linked to the other nodes, we do not gain any additional information from learning the links among these other nodes. The reason is that conditional on the node's type, it is linked to another node if and only if their types are identical. Thus, knowing the node's type and its link status with all other nodes, we can entirely pin down the rest of the network. Moreover, conditional on the node's type, its link status with respect to some node is independent of its link status with respect to another node.

It follows that the signal given by the network under perfect homophily is equivalent to a collection of $n-1$ independent signals: each signal generates a link with probability π $(1-\pi)$ conditional on the original node's type being x (y). The Bhattacharyya Coefficient for this network is thus

$$\left(\sqrt{\pi(1-\pi)}+\sqrt{(1-\pi)\pi}\right)^{n-1}$$

Since this expression is weakly lower than S, the following inequality is a necessary condition for the implementability of the first-best:

$$\left(\sqrt{4\pi(1-\pi)}\right)^{n-1} \leq (\frac{1+\lambda}{1+2\lambda})\sqrt{\frac{1-\pi}{\pi}}+(\frac{\lambda}{1+2\lambda})\sqrt{\frac{\pi}{1-\pi}}$$

By multiplying both sides of the inequality by $\sqrt{\pi/(1-\pi)}$ we can rewrite it as follows:

$$2^{n-1}\pi^{\frac{n}{2}}(1-\pi)^{\frac{n}{2}-1} \leq (\frac{1+\lambda}{1+2\lambda})+(\frac{\lambda}{1+2\lambda})(\frac{\pi}{1-\pi}) \tag{11}$$

The inequality is binding for $\pi=\frac{1}{2}$. We wish to show that there exists π^{**} sufficiently close to $\frac{1}{2}$ with the property that for every $\pi \in (\frac{1}{2},\pi^{**})$ there exists $\lambda^{**}(\pi)$ such that condition (11) is violated for every $\lambda<\lambda^{**}(\pi)$. To show this, it suffices to construct π^{**} and $\lambda^{**}(\pi)$ such that for every $\pi \in (\frac{1}{2},\pi^{**})$ and $\lambda<\lambda^{**}(\pi)$, the derivative w.r.t. π of the L.H.S of (11) is strictly higher than the corresponding derivative of the R.H.S.

The derivative w.r.t π of the L.H.S of (11) is equal to

$$2^{n-1}\pi^{\frac{n}{2}-1}(1-\pi)^{\frac{n}{2}-2}[\frac{n}{2}-\pi(n-1)] \tag{12}$$

which is positive if $\frac{1}{2}<\pi<\frac{n}{2(n-1)}$. Since the expression (12) equals 2 when $\pi=\frac{1}{2}$, is it strictly above one when π is sufficiently close to $\frac{1}{2}$.

The derivative w.r.t. π of the R.H.S. of (11) is equal to $\lambda/[(1+2\lambda)(1-\pi)^2]$, which, for all $\pi \in (\frac{1}{2},1)$, is positive and increasing in π and λ. Given $\pi \in (\frac{1}{2},\pi^{**})$, let $\lambda^{**}(\pi)$ be the solution to the equation

$$\frac{\lambda^{**}(\pi)}{1+2\lambda^{**}(\pi)} \cdot \frac{1}{(1-\pi)^2}=1$$

Hence, for any $\pi \in (\frac{1}{2},\pi^{**})$ and any $\lambda<\lambda^{**}(\pi)$, the derivative w.r.t π of the L.H.S of (11) is strictly higher than the corresponding derivative of the R.H.S. ■

Thus, an intermediate popularity gap is necessary for implementability of the first-best under the SBM. The intuition behind the case of a large popularity gap (i.e., π close to 1) is simple. For generic P and fixed n, there is an upper limit to the network's informational content, which implies a positive lower bound on the Bhattacharyya Coefficient. Moreover, this lower bound is independent of π. Therefore, a sufficiently large π induces an adverse "popularity gap" factor that overweighs whatever positive effect it may have on the informativeness factor.

The case of a low popularity gap (i.e., π close to $\frac{1}{2}$) is less obvious. In this case, the network is very uninformative about the nodes' types. For example, in the homophily case, if α is high and β is low, then with high probability the network will consist of two fully connected components, yet they will tend to be similar in size and it will be difficulty to identify the type of consumers that belong to each component. Thus, both S and $(1-\pi)/\pi$ will be close to one in the $\pi \to \frac{1}{2}$ regime, and it is not clear a priori whether the condition for implementability of the first-best will hold. However, it turns out that when π is close to $\frac{1}{2}$, the popularity-gap effect due to changing π overweighs the informativeness effect.

A simple corollary of Proposition 3 is that for every π there exists $\lambda(\pi)$ such that for every $\lambda < \lambda(\pi)$, the first-best is not implementable under any SBM with $n = 2$. This observation brings us to the role of n. The following result provides a sufficient condition for implementability of the first-best.

Proposition 4 *Fix $\pi \in (\frac{1}{2}, 1)$ and P with $p_x \neq p_y$. Then, there exists n^* such that the first-best is implementable for all SBMs (n, π, P) with $n > n^*$.*

Proof. Fix an arbitrary node i. Suppose that we were given a signal that only describes whether there is a link between i and some given node $j \neq i$. The probability of a link conditional on $t_i = x$ is $\eta_x = \pi p_x + (1-\pi)p_{xy}$, and the probability of a link conditional on $t_i = y$ is $\eta_y = \pi p_{xy} + (1-\pi)p_y$. Therefore, the Bhattacharyya Coefficient that corresponds to this signal is

$$\sqrt{\eta_x \eta_y} + \sqrt{(1-\eta_x)(1-\eta_y)}$$

Now suppose that we are given a signal that describes whether there is a link between i and *each* of the other $n-1$ nodes. Since the probability of such a link is independent across all $j \neq i$ conditional on t_i, the Bhattacharyya Coefficient that corresponds to this signal is

$$[\sqrt{\eta_x \eta_y} + \sqrt{(1-\eta_x)(1-\eta_y)}]^{n-1} \tag{13}$$

Now, observe that this signal is weakly less informative than learning the entire network w. Therefore, S is weakly below the expression (13). It follows that the following inequality is a sufficient condition for the implementability of the first-best:

$$[\sqrt{\eta_x \eta_y} + \sqrt{(1-\eta_x)(1-\eta_y)}]^{n-1} \leq (\frac{1+\lambda}{1+2\lambda})\sqrt{\frac{1-\pi}{\pi}} + (\frac{\lambda}{1+2\lambda})\sqrt{\frac{\pi}{1-\pi}} \tag{14}$$

By our assumptions on π and P, $\sqrt{\eta_x \eta_y} + \sqrt{(1-\eta_x)(1-\eta_y)} < 1$. In addition, for any π and λ that satisfy (7), the R.H.S. of (14) is bounded away from zero. Therefore, there exists n^* such that the inequality holds for every $n > n^*$. ■

Thus, for a large enough network, incentive compatibility does not constrain implementing the first-best. To see why, think of the extreme case of perfect homophily, where $\alpha = 1$ and $\beta = 0$. Then, any realized network consists of two fully connected components. When n is large, the probability that the larger component consists of x consumers is close to one. When $n \to \infty$, the network thus becomes arbitrarily informative, such that S becomes arbitrarily close to zero, and the condition for implementability of the first-best is satisfied.

To get some quantitative sense of the implications of the sufficient condition, consider the following table, which provides values of n^* for various specifications of the homophily case:

π	α	β	λ	n^*
0.6	0.1	0.05	0	1,124
0.6	0.1	0.02	0	356
0.75	0.1	0.05	0	485
0.75	0.1	0.02	0	151
0.6	0.01	0.005	0	12,060
0.6	0.01	0.002	0	3,762
0.999	0.1	0.05	0	748
0.75	0.1	0.05	0.25	232
0.75	0.1	0.02	0.25	73

This table illustrates the various forces that affect implementability of the first-best: the non-monotonic effect of π, the negative effect of low connectivity, the positive effect of strong homophily (captured by a large α/β ratio), and the positive effect of raising λ.

Up to now we assumed that the likelihood of forming links does not change as we increase the network size. Thus, the expected degree of a node was linear in n.

However, in the context of social networks, it makes sense to assume that the average number of links that a node forms grows at a slower rate than the network size. As a result, the network will become more sparse as it grows larger. In this case, it is not clear whether a larger network will be more informative than a smaller one, and therefore it is not clear whether the first-best will more easier to implement when the network is large.

To address this question, we follow the community detection literature and assume that the expected degree of a node grows *logarithmically* with n. Specifically, we assume that the connectivity matrix P depends on n, such that

$$
\begin{aligned}
p_x &= a^2 \frac{\ln(n)}{n} \\
p_{xy} &= b^2 \frac{\ln(n)}{n} \\
p_y &= c^2 \frac{\ln(n)}{n}
\end{aligned}
$$

where a, b, c are arbitrary constants. Using recent advances in the community detection literature, we derive a sufficient condition for implementability of the first-best. Specifically, we borrow existing necessary and sufficient conditions for *exact recovery* of two asymmetric "communities". By exact recovery, we mean that for a given network, there exists an algorithm that can identify the type of each node with a probability arbitrarily close to one. If exact recovery is feasible, then the network is almost perfectly informative. If node i is identified as type x (y) with probability close to one, then the probability of the observed network w is close to one conditional on $t_i = x$ (y) and close to zero conditional on $t_i = y$ (x). This implies that S is close to zero and therefore the condition for implementability of the first-best holds. For simplicity, we state the next result for the $\lambda \to 0$ limit.

Proposition 5 *In the $n \to \infty$ and $\lambda \to 0$ limit, the first-best is implementable if*

$$
\pi(a-b)^2 + (1-\pi)(c-b)^2 \geq 2 \tag{15}
$$

Proof. Let $n \to \infty$. Given the preceding paragraph, it suffices to derive a sufficient condition for exact recovery. By Abbe and Sandon (2015), such a network is exactly recoverable if and only if

$$
\max_{r\in[0,1]} \left\{ r[\pi a^2 + (1-\pi)b^2] + (1-r)[\pi b^2 + (1-\pi)c^2] - \pi a^{2r} b^{2(1-r)} - (1-\pi) b^{2r} c^{2(1-r)} \right\} \geq 1
$$

A sufficient condition for this inequality to hold is that the maximand of the L.H.S is weakly greater than one for $r = \frac{1}{2}$ - i.e., if

$$\pi(\frac{a^2+b^2}{2}) + (1-\pi)(\frac{c^2+b^2}{2}) - \pi(ab) - (1-\pi)(cb) \geq 1$$

which is equivalent to (15). ■

Note that in the homophily case, $a = c$, while in the extroversion/introversion case $b = \sqrt{ac}$. Thus, Proposition 5 implies the following.

Corollary 1 *In the $n \to \infty$ and $\lambda \to 0$ limit, the first-best is implementable in the homophily case if*

$$(a-b)^2 \geq 2$$

while in the extroversion/introversion case, the first-best is implementable if

$$(\pi a + (1-\pi)c)(\sqrt{a} - \sqrt{c})^2 \geq 2$$

Thus, when connectivity increases logarithmically with network size, a sufficient condition for implementability of the first-best for a large network is that the homophily or extroversion/introversion effects are sufficiently strong.

5 Informed Advertisers

So far, we assumed that advertisers are entirely uninformed of the realization of the network w. Relaxing this assumption raises a natural question: can the platform benefit from releasing information to the advertisers? Our first result in this section is a negative answer to this question. This finding then raises an immediate follow-up question: when advertisers can gain information about the network structure by sampling part of it, how large can this part be without destroying the platform's ability to implement the first-best?

To address the first question, suppose that an advertiser receives a signal s regarding the realization of w. Let r be the joint distribution over networks w and signals s, such that $r(s \mid w)$ is the probability that an advertiser receives the signal s conditional on the realized network being w. Conversely, let $r(w \mid s)$ be the probability that the realized network is w conditional on the signal being s. We allow signals to be correlated across advertisers conditional on w. A plausible example of a signal in this context is that the advertiser learns the subgraph induced by w over some subset of nodes. The platform does not observe the advertisers' signals.

We extend the incentive-compatibility requirement such that it needs to hold for every realization of s - that is, advertisers learn s when they report their type. In principle, because an advertiser's type now consists of both its product type and its information, one would like the pair (q, F) to condition on both. In other words, theoretically advertisers need to report both components of their type. However, because the optimal display rule is only a function of advertisers' product types, it is easy to show that the platform's ability to implement the first-best is unaffected if it also requires platforms to report their signal. Therefore, we will continue to assume that advertisers only report their product type, and this report is the only input that feeds (q, F). Then, the original IC constraints (8) are exactly the same, except that the term $\mu(w)$ is replaced with $r(w \mid s)$. We only require advertisers' IR constraint to bind ex-ante - i.e., on average across signal realizations.

It follows that in the $m \to \infty$ limit, the necessary and sufficient condition for implementability of the first-best can be written as follows. For every realization of s and every $t, t' \in \{x, y\}$,

$$\sum_{w \in W} r(w \mid s) \sum_{i \in N} q_i(t \mid w)(\rho_{t'}(i, w) - \rho_t(i, w)) \leq 0 \tag{16}$$

By Blackwell's ranking of information systems, r' is less informative than r if there is a system of conditional probabilities $(p(s \mid s'))_{s,s'}$, such that for every w, s,

$$r'(s \mid w) = \sum_{s'} p(s \mid s') r(s' \mid w)$$

The following result establishes that the platform benefits from withholding information about the network structure from advertisers. Let w^* and w_* denote the fully connected and empty networks, respectively.

Proposition 6 *(i) If the first-best is implementable under r, then it is implementable under any r' that is less informative than r. (ii) Suppose $\mu(x \mid w^*) \neq \frac{1}{2}$ or $\mu(x \mid w_*) \neq \frac{1}{2}$. Then, there exists $\lambda^* > 0$ such that if advertisers are fully informed of w (i.e., $r(w \mid w) = 1$ for every w), the first-best is not implementable for all $\lambda < \lambda^*$.*

The reason why withholding information about w from advertisers cannot harm the platform is standard - it means that IC constraints that previously held for all signals are now required to hold only on average. Part *(ii)* of the result establishes that this monotonicity result is not vacuous: giving advertisers full information about the network will prevent the platform from implementing its first-best when λ is small. This part is based on a very mild condition on the relation between the network structure and the types of individual consumers - namely, that it is impossible for both the fully connected and empty networks to induce a uniform posterior. The SBM discussed in the previous section clearly satisfies this property.

Suppose that the platform cannot prevent advertisers from getting *some* information about the network; how much information can it afford to give away? In particular, consider an SBM and assume that each advertiser get information by sampling a random subset of no more than m nodes (out of the total of n nodes in the network), and learning the subgraph over these m nodes that is induced by w. Recall that w is realized according to a given SBM. Hence, the Bhattacharyya Coefficient can be defined for any subgraph of w consisting of k nodes, $k = 1, ..., n$ (where the connectivity matrix is fixed). Denote this coefficient by $S(k)$.

Proposition 7 *Suppose each advertiser is informed of the subgraph induced by w over a random subset of at most m nodes. If*

$$S(n-m) \leq \frac{1}{\sqrt{\pi(1-\pi)}}\left[\frac{\lambda+(1-\pi)}{1+2\lambda} - (\frac{m}{n-m})(\frac{\sqrt{2}-1}{2})\right] \tag{17}$$

then the first-best is implementable.

Proof. Suppose an advertiser learns the subgraph induced by w over some subset of nodes N_1 (the size of which is n_1). We can represent w as a triple (g_1, g_2, h), where g_1 is the subgraph that the advertiser learns, g_2 is the subgraph induced by w over the remaining set of nodes $N_2 = N - N_1$ (the size of which is n_2), and h consists of all links between a node in N_1 and a node in N_2. Because w is generated by an SBM and g_1 and g_2 are defined over disjoint sets of nodes, g_1 and g_2 are independently distributed.

The necessary and sufficient condition for implementability of the first-best is that for every signal g_1,

$$\sum_{g_2,h}\mu(g_2,h \mid g_1)\sum_{i\in N}\sqrt{\mu_i(x \mid g_1,g_2,h)\mu_i(y \mid g_1,g_2,h)} \leq \tag{18}$$

$$\sum_{g_2,h}\mu(g_2,h \mid g_1)\sum_{i\in N}\left[\frac{1+\lambda}{1+2\lambda}\mu_i(y \mid g_1,g_2,h) + (\frac{\lambda}{1+2\lambda})\mu_i(x \mid g_1,g_2,h)\right]$$

and

$$\sum_{g_2,h}\mu(g_2,h \mid g_1)\sum_{i\in N}\sqrt{\mu_i(x\mid g_1,g_2,h)\mu_i(y\mid g_1,g_2,h)}\leq \tag{19}$$
$$\sum_{g_2,h}\mu(g_2,h \mid g_1)\sum_{i\in N}\left[\frac{1+\lambda}{1+2\lambda}\mu_i(x\mid g_1,g_2,h)+\frac{\lambda}{1+2\lambda}\mu_i(y\mid g_1,g_2,h)\right]$$

By construction, g_1 and g_2 are independent, such that we can write

$$\mu(g_2,h\mid g_1)=\mu(g_2)\mu(h\mid g_1,g_2)$$

and observe that

$$\mu_i(x\mid g_1,g_2)=\sum_h \mu(h\mid g_1,g_2)\mu_i(x\mid g_1,g_2,h)$$

Applying the Cauchy-Schwartz inequality, we obtain

$$\sqrt{\mu_i(x\mid g_1,g_2)\mu_i(y\mid g_1,g_2)}\geq\sum_h\mu(h\mid g_1,g_2)\sqrt{\mu_i(x\mid g_1,g_2,h)\mu_i(y\mid g_1,g_2,h)}$$

It follows that inequalities (18)-(19) are implied by the following, simpler inequalities:

$$\sum_{i\in N}\left[\sum_{g_2}\mu(g_2)\sqrt{\mu_i(x\mid g_1,g_2)\mu_i(y\mid g_1,g_2)}-\frac{1+\lambda}{1+2\lambda}\mu_i(y\mid g_1)-\frac{\lambda}{1+2\lambda}\mu_i(x\mid g_1)\right]\ \leq\ 0$$
$$\sum_{i\in N}\left[\sum_{g_2}\mu(g_2)\sqrt{\mu_i(x\mid g_1,g_2)\mu_i(y\mid g_1,g_2)}-\frac{1+\lambda}{1+2\lambda}\mu_i(x\mid g_1)-\frac{\lambda}{1+2\lambda}\mu_i(y\mid g_1)\right]\ \leq\ 0$$

Consider first the top inequality. We can break the summation over $i\in N$ into two summations over N_1 and N_2. Because g_1 and g_2 are independent, for every $i\in N_1$ we can write $\mu_i(x\mid g_1,g_2)=\mu_i(x\mid g_1)$. Similarly, for every $i\in N_2$ we can write $\mu_i(x\mid g_1,g_2)=\mu_i(x\mid g_2)$ and $\mu_i(x\mid g_1)=\mu_i(x)=\pi$. It follows that the inequality can be rewritten as

$$\begin{aligned}
&\sum_{i\in N_2}\left[\sum_{g_2}\left(\mu(g_2)\sqrt{\mu_i(x\mid g_2)\mu_i(y\mid g_2)}-(\frac{1+\lambda}{1+2\lambda})\mu_i(y\mid g_1)-(\frac{\lambda}{1+2\lambda})\mu_i(x\mid g_1)\right)\right]+\\
&\sum_{i\in N_1}\left[\sum_{g_2}\left(\mu(g_2)\sqrt{\mu_i(x\mid g_1)\mu_i(y\mid g_1)}-(\frac{1+\lambda}{1+2\lambda})\mu_i(y\mid g_1)-(\frac{\lambda}{1+2\lambda})\mu_i(x\mid g_1)\right)\right]\\
\leq\ & 0
\end{aligned}$$

The top sum can be simplified into

$$n_2S(n_2)\sqrt{\pi(1-\pi)}-n_2(\frac{1+\lambda}{1+2\lambda})(1-\pi)-n_2(\frac{\lambda}{1+2\lambda})\pi$$

while the bottom sum can be grouped together as

$$\begin{aligned}
&\sum_{i\in N_1}\left[\sqrt{\mu_i(x\mid g_1)\mu_i(y\mid g_1)}-(\frac{1+\lambda}{1+2\lambda})\mu_i(y\mid g_1)-(\frac{\lambda}{1+2\lambda})\mu_i(x\mid g_1)\right]\\
\leq\ & n_1\cdot\max_{a\in\{0,1\}}\max_{d\in[0,1]}\left[\sqrt{d(1-d)}-ad-(1-a)d\right]\\
=\ & n_1\cdot\frac{\sqrt{2}-1}{2}
\end{aligned}$$

Plugging this term and exploiting the assumption that $\pi>\frac{1}{2}$, we can now obtain the following sufficient condition for implementability of the first-best:

$$n_2\left[S(n_2)\sqrt{\pi(1-\pi)}-\frac{\lambda\pi}{1+2\lambda}-\frac{(1+\lambda)(1-\pi)}{1+2\lambda}\right]+n_1\frac{\sqrt{2}-1}{2}\leq 0 \qquad (20)$$

Substituting m for n_1 and $n-m$ for n_2 yields the desired condition. ■

Inequality (17) is a sufficient condition for implementability of the first-best. When π and the connectivity matrix that define the SBM are fixed, the inequality is stated entirely in terms of m and n. In principle, it is also possible to calculate $S(n-m)$, and because (17) is a sufficient condition, we can use upper bounds on $S(k)$ derived in Section 4 to get a closed-form upper bound on m, such that the first-best is implementable for any value of m below that bound. Finally, the comparative statics w.r.t m is consistent with our previous results. When m increases, the R.H.S of (17) clearly goes down, whereas $S(n-m)$ goes up, because a smaller network is a less informative signal. Thus, a larger m makes is more difficult to satisfy the sufficient condition.

6 Extensions

In this section we briefly discuss two natural extensions of our model.

Additional consumer observables

As mentioned before, the basic results of Section 3 do not rely on the interpretation of w as a network - it can be any signal about consumer types. In particular, the description of w may consist of the realized network as well as personal information about consumers (demographic characteristics, actions they took within the confines of the social network that indicate their preferences, etc.). The following is an example of how this extension may change the analysis of the SBM specification in Section 4.

Suppose that a consumer $i \in N$ is characterized by a pair (t_i, a_i), where $t_i \in \{x, y\}$ is the usual preference type and $a_i \in \{1, 2\}$ is a demographic characteristic. The ex-ante probability that $a_i = 1$ (for any i), denoted γ, is independent of t_i. As before, the platform does not observe t_i. However, we assume that the platform does observe a_i. Thus, w consists of the realized network and the profile $(a_1, ..., a_n)$. Suppose that if $a_i \neq a_j$, then consumers i and j form a link with probability *zero*. The probability of a link between consumers i and j with $a_i = a_j = a$ is a function of t_i and t_j, given by a symmetric 2×2 connectivity matrix P^a, as before.

We now provide a characterization of the Bhattacharyya Coefficient that is induced by this process. When the platform learns the profile $(a_1, ..., a_n)$, it can partition N into two sets, N_1 and N_2, such that $a_i = k$ for every $i \in N_k$. By assumption, the two subgraphs over N_1 and N_2 induced by the network, denoted w_1 and w_2, are necessarily mutually isolated. The absence of a link between nodes that belong to different sets conveys no information, given that we know the realization $(a_1, ..., a_n)$. Therefore, w_1 and w_2 are independently drawn conditional on the realization of $(a_1, ..., a_n)$. The Bhattacharyya Coefficient of a signal that consists of the two graphs over N_1 and N_2 is thus $S(1, n_1) \cdot S(2, n_2)$, where $S(k, n_k)$ is the Bhattacharyya Coefficient of an SBM defined by (n_k, π, P^k). To obtain the Bhattacharyya Coefficient of the entire process, we need to average out over all possible realizations of N_1 and N_2:

$$S = \sum_{r=1}^{n} \binom{n}{r} \gamma^r (1-\gamma)^{n-r} S(1, r) S(2, n-r)$$

We can now plug this expression for S into necessary or sufficient conditions for implementability of the first-best.

Many product types

Throughout this paper, we assumed that there are only two product types, x and y.

Now suppose that there are $K > 2$ product types, denoted $x_1, ..., x_K$. Consider the case in which $q_i(x_k \mid w) > 0$ for all $i \in N$, $w \in W$ and $k = 1, ..., K$. Then, it is straightforward to show that a necessary and sufficient condition for implementability of the first-best is that for every pair of types x_k and x_m,

$$S(k, m) \leq (\frac{1+\lambda}{1+2\lambda})\sqrt{\frac{\mu(x_k)}{\mu(x_m)}} + (\frac{\lambda}{1+2\lambda})\sqrt{\frac{\mu(x_m)}{\mu(x_k)}}$$

where $\mu(x_k)$ is the ex-ante probability that a consumer wants x_k, and $S(k, m)$ is the Bhattacharyya Coefficient of the conditional distributions $(\mu(w \mid x_k))_{w \in W}$ and $(\mu(w \mid x_m))_{w \in W}$. This condition is an immediate extension of the condition for two product types.

7 Conclusion

We presented a modeling framework that sheds light on some of the forces that shape the optimal design of incentive-compatible advertising on social networks. Our results illuminate how the distribution over consumer preferences and the process that generates the social network affect the incentive issues that the platform faces when designing personalized display rules and setting advertising fees. The results also demonstrate the non-trivial effects of consumers' attentiveness to advertising and their frequency of repeat purchases. Finally, the analysis highlights - and to some extent quantifies - the importance of keeping advertisers uninformed of the structure of the social network. Perhaps the most intriguing aspect of our model is the connection to the community detection problem. Given that the latter is an active research area in Network Science, we hope that future results in this literature will generate additional insights into the question of incentive-compatible advertising on a social network.

Appendix: Proofs

Proposition 1

From (7), it follows that (5) characterizes the first-best display policy. Plugging this expression for $q_i(t \mid w)$ into the $IC(x, y)$ constraint (8) yields the following inequality

$$\sum_{i \in N} \sum_{w \in W} \mu(w) \cdot \sqrt{\mu_i(x \mid w)\mu_i(y \mid w)} \leq \sum_{i \in N} \sum_{w \in W} \mu(w) \cdot [(\frac{1+\lambda}{1+2\lambda})\mu_i(y \mid w) + (\frac{\lambda}{1+2\lambda})\mu_i(x \mid w)]$$

Note that $\mu(w)\mu_i(t \mid w) = \mu_i(t, w)$ and $\sum_{w \in W} \mu_i(x, w) = \pi$. The above inequality can thus be rewritten as

$$\sum_{i\in N}\sum_{w\in W}\sqrt{\mu_i(x,w)\mu_i(y,w)} \leq n(\frac{1+\lambda}{1+2\lambda})(1-\pi)+n(\frac{\lambda}{1+2\lambda})\pi \tag{21}$$

Because $\mu_i(x,w)=\pi\mu_i(w\mid x)$ and $\mu_i(y,w)=(1-\pi)\mu_i(w\mid y)$, we can express (21) as the following inequality,

$$\sum_{i\in N}\sum_{w\in W}\sqrt{\mu_i(w\mid x)\mu_i(w\mid y)} \leq n(\frac{1+\lambda}{1+2\lambda})\sqrt{\frac{1-\pi}{\pi}}+n(\frac{\lambda}{1+2\lambda})\sqrt{\frac{\pi}{1-\pi}}$$

By the ex-ante symmetry of nodes, the L.H.S. of the above inequality is simply nS, so that this inequality reduces to

$$S \leq \frac{1+\lambda}{1+2\lambda}\sqrt{\frac{1-\pi}{\pi}}+\frac{\lambda}{1+2\lambda}\sqrt{\frac{\pi}{1-\pi}} \tag{22}$$

If we carry out a similar exercise for $IC(y,x)$, we obtain the inequality

$$S \leq \frac{1+\lambda}{1+2\lambda}\sqrt{\frac{\pi}{1-\pi}}+\frac{\lambda}{1+2\lambda}\sqrt{\frac{1-\pi}{\pi}}$$

By assumption, $\pi \geq \frac{1}{2}$. And since $1+\lambda > \lambda$, the only inequality that matters is (22), which is precisely the condition (9).

Proposition 6

(i) The proof is entirely rudimentary and standard. Nevertheless, we give it for completeness. By assumption, inequality (16) holds for every s. Using the definition of Blackwell informativeness, we can write

$$\begin{aligned} r'(w \mid s) &= \frac{\mu(w)}{r'(s)}r'(s\mid w) = \frac{\mu(w)}{r'(s)}\sum_{s'}p(s\mid s')r(s'\mid w) \\ &= \frac{\mu(w)}{r'(s)}\sum_{s'}p(s\mid s')\frac{r(s')r(w\mid s')}{\mu(w)} = \sum_{s'}\frac{p(s\mid s')r(s')}{r'(s)}r(w\mid s') \end{aligned}$$

where $r(s')$ is the ex-ante probability of the signal s' under r, and $r'(s)$ is the ex-ante probability of the signal s under r'. Now, elaborate the term

$$\frac{p(s\mid s')r(s')}{r'(s)} = \frac{\sum_w \mu(w)p(s\mid s')r(s'\mid w)}{\sum_{s''}\sum_w \mu(w)p(s\mid s'')r(s''\mid w)}$$

We can easily see that this term is between 0 and 1, and that

$$\sum_{s'} \frac{p(s \mid s')r(s')}{r'(s)} = 1$$

It follows that for every s, $r'(w \mid s)$ is some convex combination of $(r(w' \mid s))_{w'}$. Therefore, given that under r, (16) holds for every s, it must hold under r' as well.

(ii) Suppose that advertisers are fully informed of the realization of w. Then, the necessary and sufficient conditions for implementability of the first-best are that for every w,

$$\begin{aligned}
\sum_{i\in N} \sqrt{\mu_i(x \mid w)\mu_i(y \mid w)} &\leq \sum_{i\in N} \left[\frac{1+\lambda}{1+2\lambda}\mu_i(y \mid w) + \frac{\lambda}{1+2\lambda}\mu_i(x \mid w)\right] \\
\sum_{i\in N} \sqrt{\mu_i(x \mid w)\mu_i(y \mid w)} &\leq \sum_{i\in N} \left[\frac{1+\lambda}{1+2\lambda}\mu_i(x \mid w) + \frac{\lambda}{1+2\lambda}\mu_i(y \mid w)\right]
\end{aligned}$$

Let $w \in \{w_*, w^*\}$. Then, w is a symmetric signal - i.e., $\mu_i(x \mid w)$ is the same for all $i \in N$, such that we can remove the subscript i and the summation over i from both inequalities. The inequalities then reduce to

$$1 \leq (\frac{1+\lambda}{1+2\lambda})\sqrt{\frac{\mu(y \mid w)}{\mu(x \mid w)}} + (\frac{\lambda}{1+2\lambda})\sqrt{\frac{\mu(y \mid w)}{\mu(x \mid w)}} \tag{23}$$

$$1 \leq (\frac{1+\lambda}{1+2\lambda})\sqrt{\frac{\mu(x \mid w)}{\mu(y \mid w)}} + (\frac{\lambda}{1+2\lambda})\sqrt{\frac{\mu(x \mid w)}{\mu(y \mid w)}} \tag{24}$$

Because $\mu(x \mid w^*) \neq \frac{1}{2}$ or $\mu(x \mid w_*) \neq \frac{1}{2}$, either $\mu(x \mid w) > \mu(y \mid w)$ or $\mu(x \mid w) > \mu(y \mid w)$. Assume w.l.o.g that $\mu(x \mid w^*) > \mu(y \mid w^*)$. Since inequality (24) is violated for $\lambda = 0$, there exists $\lambda^* > 0$ such that this inequality would also be violated for all $\lambda < \lambda^*$.

References

[1] Abbe, Emmanuel and Colin Sandon (2015): "Community detection in general stochastic block models: fundamental limits and efficient recovery algorithms," Working Paper, Princeton University (arXiv:1503.00609v2).

[2] Athey, Susan and Joshua S. Gans (2010): "The impact of targeting technology on advertising markets and media competition," *American Economic Review: Papers & Proceedings* 100(2), 608-613.

[3] Basu, Ayanendranath, Hiroyuki Shioya and Chanseok Park (2011): *Statistical inference: the minimum distance approach.* CRC Press.

[4] Bergemann, Dirk and Alessandro Bonatti (2011): "Targeting in advertising markets: implications for offline verus online media," *Rand Journal of Economics* 42(3), 417-443.

[5] Bloch, Francis (2015): "Targeting and pricing in social networks," in *The Oxford Handbook of the Economics of Networks* (Y. Bramoulle, A. Galeotti and B. Rogers eds.), Oxford University Press.

[6] Campbell, James (2015): "Localized price promotions as a quality signal in a publicly observable network," *Quantitative Marketing and Economics* 13(1), 27-57.

[7] Eliaz, Kfir and Ran Spiegler (2015): "Search design and broad matching," forthcoming in *American Economic Review.*

[8] Galeotti, Andrea and Sanjeev Goyal (2012): "Network multipliers: the optimality of targeting neighbors," *Review of Network Economics* 11(3), 1446-9022.

[9] Iyer, Ganesh, David Soberman and J. Miguel Villas-Boas (2005): "The targeting of advertising," *Management Science* 24(3), 461-476.

[10] Johnson, Justin P. (2013): "Targeted advertising and advertising avoidance," *Rand Journal of Economics* 44(1), 128-144.

[11] Kandell, Denise B. (1978): "Homophily, selection and socialization in adolescent friendships," *American Journal of Sociology* 84(2), 427–36.

[12] Mossel, Elchanan. Joe Neeman and Allan Sly (2012): "Stochastic block models and reconstruction," mimeo.

[13] Theodoridis, Sergios and Konstantinos Koutroumbas (2008): *Pattern Recognition.* Academic Press.

[14] Zubcsek, Peter Pal and Miklos Sarvary (2011): "Advertising to a social network," *Quantitative Marketing and Economics* 9(1), 71-107.