Incentive-Compatible Advertising on a Social Network*

Kfir Eliaz and Ran Spiegler†

May 14, 2017

Abstract

A platform that operates a social network allows firms to post display ads to consumers on the network. The network structure is correlated with the profile of members' preference types. The platform's policy consists of a personalized, stationary ad-display rule and an advertising fee (which the platform charges from firms as a function of the consumer type they request to target). 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 propensity for repeat 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 a "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 Yair Antler, Francis Bloch, Olivier Compte, Andrea Galeotti, Nikita Roketskyi, Omer Tamuz, Neil Thakral and numerous seminar audiencces 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 targeting opportunities for advertisers. 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 of consumer preference types, which in turn helps the platform increase its advertising revenues. In this paper, we study targeted advertising when the platform operates a *social network*. The platform's primary function is not commercial but social: consumers use it to cultivate relationships with other network members. However, consumers' social activity generates valuable information for advertisers. In particular, a consumer's exact network location is indicative of his tastes. For instance, if consumers exhibit *homophily* - i.e., they associate with like-minded individuals (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 questions such as: How can the platform use the information inherent in the network's structure to allocate display-ad slots in a way that maximizes advertisers' surplus? What are the incentive constraints that the platform faces when advertisers have private information regarding the quality of their match with various consumer types? 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 and propensity for repeat purchases, or the extent to which advertisers are informed of the network structure?

In our model, there is a group of n consumers, each coming in one of two (private) preference types. A type can describe whether the consumer is interested in "healthy food", whether he likes "highbrow" movies, whether he enjoys outdoor recreational activities, etc. Consumers are linked by a social network, the structure of which is a (commonly known) stochastic function of consumers' types. The platform observes the network and updates its belief regarding consumers' types. It then enables advertisers (a.k.a firms) to post personalized display ads. Each firm is characterized by the quality of its match with each consumer type - defined as the probability of transaction conditional on the consumer's exposure to the firm's ad. Ex-ante, each firm communicates to the platform the type of consumers it would have liked to target. (In the basic model, firms are uninformed about the network structure.) Thus, ads are classified into "types" according to the targeting request that accompanies them. Advertiser-platform communication of this kind exists in reality: Facebook operates an interface that allows advertisers to indicate the preference groups they wish to target.[1]

We assume that consumers' exposure to ads is governed by a personalized, stationary display rule that the platform designs. Specifically, the platform tailors a mixture of ad types to each consumer - based on its updated belief regarding his type - such that the ad he is exposed to at any period is drawn independently according to this mixture. (We think of ads as "billboards" and interpret transactions as occurring "offline", unmonitored by the platform.) As soon as the consumer transacts with an advertiser, he switches to a "satiation" mental state in which he is inattentive to ads, and he switches back to the attentive state of mind with some constant per-period probability that captures the propensity for repeat purchases. Thus, thanks to the simplifying assumption of stationary display rules, we can depict the consumer's experience at the social network as a personal two-state Markov process, where certain transition probabilities are determined by the platform's personalized display rule.

The platform's objective is to maximize total advertisers' surplus - defined as their long-run number of transactions per period, calculated according to consumers' personal Markov processes - and to extract it by means of advertising fees. We describe the fee as a constant price per period that depends on the firm's targeting request. By the assumption of risk neutrality, the fee can equivalently (and more realistically) be described as a price-per-display, which may also be a function of the realized network. The platform's policy thus consists of its display rule and fee schedule.

A key observation in this model is that because of the platform's uncertainty regarding consumers' types and their mental state at any given period, its optimal display rule may be *interior* - i.e., it may expose individual consumers to *both* ad types. This turns out to imply a motive for advertisers to *strategize* their targeting request. Before explicating this incentive issue, we first need to understand why the platform relies on advertisers' requests in the first place. In principle, it could examine each advertiser's product and figure out the quality of its match with each consumer type. However, this type of monitoring has a cost that the platform can avoid by decentralizing the ad-classification task. Furthermore, in many cases advertisers have private information regarding the type of consumers who are attracted to their product, thanks to prior market research. For instance, certain food items (granola bars, artificially sweetened products) are not easy to classify a priori in terms of their appeal to "health-conscious" consumers. Likewise, the defining lines of "highbrow" movies or holiday resorts that fit "outdoorsy" tourists are quite blurred. In these cases, market studies are likely

[1] See https://www.facebook.com/business/products/ads/ad-targeting/

to reveal information that the platform lacks. It is implausible for the platform to replicate such studies in the myriad industries it interacts with.

What is the force that could lead an advertiser to strategize its targeting request? The key insight is that optimal display rules approximately minimize the amount of time that it takes a non-satiated consumer to transact. As a result, the optimal probability that a firm is displayed to a particular consumer is approximately proportional to the *square root* of the posterior probability (given the realized network) that the firm's product fits the consumer's type. In contrast, the advertising fee that fully extracts a firm's surplus is proportional to the *prior probability* that consumers like its product. This discrepancy ends up discriminating against products with mass appeal, and it gives firms an incentive to target the minority consumer group: the reduction in display frequency will be more than compensated for by the reduced advertising fee. Indeed, this trade-off between the value of exposure to a target audience and the cost of reaching it is a dilemma that advertisers often face, and which various intermediaries offer to solve by searching for the target audience that gives the "best bang for the buck".[2] To achieve this objective, these intermediaries may divert their client's ad to a less-than-ideal audience because it may be significantly cheaper.

We begin our analysis by posing the following question: Under what conditions on the environment's primitives can the platform design an incentive-compatible policy that maximizes and fully extracts advertisers' surplus? Our interest in this question is two-fold. First, it serves as a useful theoretical benchmark for the platform's design problem. Second, and perhaps more interestingly, it can be interpreted in the spirit of "welfare theorems" in the competitive-equilibrium literature. We can think of the platform's policy as a market institution for allocating advertisers to consumers' limited attention, which is the scarce resource in this environment. The full-surplus-extraction requirement is essentially a zero-profit condition that captures competitive behavior among advertisers. Our question then becomes: Can an efficient allocation of advertisers to network users be supported by a competitive market?

Our basic result is a necessary and sufficient condition for the implementability of the platform's objective (assuming that exogenous parameters are such that the optimal display rule is interior for *all* consumers and network realizations - otherwise, our condition is merely sufficient). The condition is a simple inequality that incor-

[2]For instance, AdEspresso.com, is a company that offers to help small businesses launch advertising campaigns on Facebook. On their website they write, "The audience you choose will directly affect how much you're paying for Facebook Ads ... if your perfect audience is just more expensive, that's just the way it goes." The company Brandnetworks (bn.co) offers algorithmic ad management that maximizes the cost and performance of ads in real-time.

porates two quantities: (i) the extent to which the social network is informative of consumers' types - measured by the *Bhattacharyya Coefficient* of similarity between the distributions over realized networks conditional on the two possible type realizations of a single consumer; and (ii) the ratio between the ex-ante probabilities of the two consumer types. The inequality is *easier to satisfy* when the network becomes *more informative*, when consumers are *less attentive* to ads, when the gap between high- and low-quality match probabilities is *smaller*, and when repeat purchases are *more* frequent. The result greatly simplifies applications, because we can often rank network-generating processes in terms of their informativeness. (It also answers the other comparative-statics questions we posed in the opening paragraphs.) However, the quantities (i) and (ii) above are *not* independent, because the network's informativeness varies with the consumer-type distribution. This interdependence leads to a few surprising conclusions.

The main part of our analysis is an application of the basic result to a familiar model of random networks known as the "*stochastic block model*" (SBM). This model has been extensively studied in the Computer Science literature (Mossel et al. (2012), Abbe and Sandon (2015)). It is parameterized by the size of the network n, the distribution over consumer types and a "connectivity matrix" that specifies the (independent) probability of a link between any pair of consumers as a function of their types. The SBM subsumes two natural models of network formation: the "homophily" case where identical types are more likely to form a link, and an "extroversion/introversion" case where the probability of a link between two consumers is the product of their individual propensities to form a link.

Our analysis of the SBM focuses on the parameter regime of low repeat-purchase propensity and large match-quality gap, where the optimal display rule is interior and its implementation is more difficult. We show that the first-best is not implementable if the consumer type distribution is either too asymmetric or too uniform. We also use the SBM 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 d nodes; how large can d be without destroying the implementability of the platform's objective? We present a sufficient condition that quantifies the upper bound on d in terms of the consumer type distribution and the Bhattacharyya Coefficient induced by the SBM with $n - d$ nodes.

A natural question for the SBM is whether a larger network makes it easier for the platform to implement its objective. For a fixed connectivity matrix, implementability is secured if n is sufficiently large. This is a straightforward "law of large numbers" effect: the network becomes arbitrarily informative in the $n \to \infty$ limit. The literature on the so-called "community detection" problem has addressed a more challenging 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 platform's objective. For homophily or extroversion/introversion, the condition translates into a lower bound on the strength of these effects. Thus, our analysis uncovers a connection between the community-detection problem in Network Science and the economic question of incentivizing targeted advertising on social networks.

Related literature
This paper belongs to a research agenda that explores novel incentive issues in modern platforms. Our earlier exercise in this vein, Eliaz and Spiegler (2015), studied an environment in which consumers submit noisy queries to a "search platform", which responds 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 search pools and extracting their 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 more elaborate form. New substantive and technical questions arise from the social-network context.

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 et al. (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). Our model entirely abstracts from this important consideration.

Our model is also related to the question of how a monopolistic data provider should price chunks of information ("cookies") about individual consumers' characteristics. Bergemann and Bonatti (2015) proposed a model in which a single data provider sells information to firms about the potential value of matches with various consumers. Each firm chooses the set of consumers whose match value it wants to know, and how much to invest in advertising to each consumer (this investment affects the probability of realizing the match value). The authors characterize the equilibrium of this market and show when advertisers would want to purchase information about consumers with high versus low match values. Our paper is complementary to theirs. While Bergemann and Bonatti study the sale of information about specific consumers to advertisers, we focus instead on the sale of personalized ad slots. Indeed, their model provides an additional justification for our assumption that advertisers are better informed than the platform about the quality of a match with various consumer types, because they acquire this information from a third party.

2 A Model

Let $N = \{1, ..., n\}$ be a set of consumers and let $T = \{x, y\}$ be a set of consumer types. We consider a stationary environment in which a consumer of type t is in one of two states: a "demand state" D_t in which the consumer buys a product with positive probability when he is exposed to an ad for it (we describe below the ad-display and purchase processes) and a "satiation state" S_t in which the consumer is not interested in consumption. The consumer switches from his demand state to his satiation state as soon as he buys a product. When the consumer is in his satiation state, he switches back to his demand state with independent per-period probability ε. This parameter captures consumers' propensity for repeat purchases.

The set of products offered by advertisers can be partitioned into two types, also labeled x and y. The probability that a consumer buys a product conditional on being exposed to its ad (as well as being in his demand state) is θ_H when the product's type coincides with his own type, and θ_L when it does not, where $\theta_L < \theta_H$. The parameters θ_L, θ_H also reflect consumers' general attention to ads: raising both by a common factor captures greater attentiveness. We say that an advertiser is of type t if it offers a type t product. There are m advertisers of each type (our analysis will focus on the case of large m). Each advertiser can costlessly supply any amount of its product. If a consumer acquires a product from an advertiser, the advertiser earns a fixed payoff of 1 (we abstract from product pricing).[3]

Consumers and advertisers are brought together by a platform that operates a social

[3] 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.

network - i.e., it enables consumers to form social links with each other. Whether a pair 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 or 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 $t_i = x$, and let $\mu(w) = \sum_x \mu_i(x,w)$ be the ex-ante probability that the realized network is w. Given some network w, the conditional probability that $t_i = x$ is denoted $\mu_i(x|w)$. Likewise, $\mu(w|t_i)$ describes 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, $\mu_i(x|w) = \mu_j(x|w)$. Finally, denote $\mu(x) = \pi \geq \frac{1}{2}$.

We assume that the platform observes the network, but does not directly observe the types of consumers and advertisers. Consumers' and advertisers' types are their private information. Advertisers do not observe the realized social network - we relax this assumption in Section 5.

Example 1: 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 $\mu_i(w|t_i)$. 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. □

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 stationary random process we will describe momentarily and displays it to the consumer in the form of an advertising banner. Each ad expires at the end of the period and a new one is displayed in the next period. We think of ads as "billboards": transactions between consumers and advertisers take place "offline", and therefore the platform cannot monitor them. In particular, there is no notion of "clicking" on display ads.

The display of ads is governed by a personalized, stationary rule that the platform commits to ex-ante. Formally, $q_i(t|w)$ 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, $q_i(y|w) \equiv 1 - q_i(x|w)$. 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 $q_i(t|w)/m$. We refer to $q(w) = (q_i(x|w))_{i \in N}$ 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*.

Given that the consumer cycles between his two mental states indefinitely, the assumption of stationary display rules means that the consumer's behavior over time obeys a two-state Markov process. Formally, given a network w and a display rule $q(w)$, the transition probabilities between the mental states of consumer i of type t are given by the following matrix:

$$\begin{array}{ccc} & D_t & S_t \\ D_t & 1 - \theta_H q_i(t|w) - \theta_L(1 - q_i(t|w)) & \theta_H q_i(t|w) + \theta_L(1 - q_i(t|w)) \\ S_t & \varepsilon & 1 - \varepsilon \end{array} \tag{1}$$

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

$$\rho_i(t|w) \equiv \frac{\mu_i(t|w)\varepsilon}{q_i(t|w)(\theta_H - \theta_L) + \theta_L + \varepsilon} \tag{2}$$

As will be shown shortly, this formula leads to a simple expression for the long-run average number of transactions for each consumer.

2.1 Discussion

Consumer and advertiser types. We envision the consumer as a vector of unobservable personality attributes. (We introduce observed characteristics such as age, education or hobbies in Section 6.) For simplicity, our model assumes that the possible realizations of this collection of attributes can be partitioned into two groups, x and y, such that every product that is offered in the market is more appealing to one of the two groups. Thus, we interpret advertisers of type x (y) as offering a *variety* of products, which all share the feature that they are more appealing to x (y) consumers.

Per-period fee vs. price per-display. Our analysis will remain unchanged if we assume that instead of charging a per-period fee, the platform charges a *price-per-display.*

Likewise, we could allow prices to be a function of the realized network w, without any effect on our analysis. This is because advertisers in our model are risk neutral and care only about the expected number of transactions and the expected payment. It is therefore convenient analytically to assume lump-sum transfers, even if this may appear unrealistic when taken literally.

The stationarity assumption. Stationary in our model has exogenous and endogenous aspects. The former arises from our assumption that ads are "billboards"; the platform cannot monitor whether consumers pay attention to ads and whether they transact with firms. Therefore, it cannot learn anything about consumers' types after the network is realized. Even in this case, stationary display rules (which is the endogenous aspect) carry loss of generality. If we relaxed this assumption and allowed the platform's display rule to follow some Markov process with an arbitrary number of states K, the consumer's behavior over time would obey a $2K$-state Markov process, and therefore the long-run average number of transactions would be hard to characterize.

Suppose that ads are not billboards, such that the platform can partially monitor whether consumers notice them - e.g. through *clicks* (and let us retain the rather realistic assumption that the platform does not monitor transactions). If clicks are uncorrelated with consumers' types, then exogenous stationarity continues to hold, and pricing displays is equivalent to pricing clicks. Therefore, we can think of our stationarity and pricing assumptions as reasonable approximations to situations in which clicks convey little information about consumer types (e.g., because an ad's ability to draw attention is rather independent of the consumer's personality type).

Ultimately, we think of stationarity as a plausible simplifying assumption in what is (to our knowledge) the first model of incentive-compatible targeted advertising on social networks. Examining which of our insights survive the extension to non-stationary environments is left for future research.

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.

The results in this section only rely on the view of $\mu_i(\cdot|t)$ as a Blackwell information system. The fact that w is a social network plays no role, and the same analysis would hold if w were an arbitrary joint signal of $(t_1, ..., t_n)$. 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 (e.g., a communication platform such as e-mail or chats). More importantly, in Section 4 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.

3.1 Optimal Policies

The platform's objective is to find a policy (q, F) that maximizes expected profits. 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 multiply the invariant probability that the consumer is in his demand state by the probability that he transacts conditional on being in this state. Then, we sum over all consumers. It follows that the expected number of transactions per period with advertisers of type x is

$$U_x(q) \equiv \sum_{i \in N} \sum_{w \in W} \mu(w) q_i(x|w)[\theta_H \rho_i(x|w) + \theta_L \rho_i(y|w)] \tag{3}$$

Similarly, the expected number of transactions per period with advertisers of type y is

$$U_y(q) \equiv \sum_{i \in N} \sum_{w \in W} \mu(w) q_i(y|w)[\theta_H \rho_i(y|w) + \theta_L \rho_i(x|w)] \tag{4}$$

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

$$U_x(q) + U_y(q) \tag{5}$$

We refer to $q^*(w)$ as the optimal display rule for the network w. The optimal fee that the platform charges advertisers of type t, denoted F_t^*, fully extracts the maximal surplus of these advertisers - i.e., $F_t^* = U_t(q^*(w))$.

Note that for a fixed μ, if consumers are sufficiently inattentive to ads in the sense that both θ_H and θ_L are sufficiently close to zero, the optimal display rule is generically a corner solution: $q_i^*(t|w) = 1$ if $\mu_i(t|w) > \frac{1}{2}$. When $q_i^*(w)$ is interior, first-order conditions imply

$$\frac{\rho_i(x|w)}{\rho_i(y|w)} = \sqrt{\frac{\mu_i(x|w)}{\mu_i(y|w)}} \tag{6}$$

Substituting the R.H.S. of (2) for $\rho_i(x|w)$ allows us to solve explicitly for $q_i^*(x|w)$:

$$q_i^*(x|w) = \frac{\lambda_H\sqrt{\mu_i(x|w)} - \lambda_L\sqrt{\mu_i(y|w)}}{\sqrt{\mu_i(x|w)} + \sqrt{\mu_i(y|w)}} \tag{7}$$

where

$$\lambda_H \equiv \frac{\theta_H + \varepsilon}{(\theta_H + \varepsilon) - (\theta_L + \varepsilon)} \text{ and } \lambda_L \equiv \frac{\theta_L + \varepsilon}{(\theta_H + \varepsilon) - (\theta_L + \varepsilon)}$$

(note that $\lambda_H - \lambda_L = 1$). As $\lambda_L \to 0$,

$$q_i^*(x|w) \to \frac{\sqrt{\mu_i(x|w)}}{\sqrt{\mu_i(x|w)} + \sqrt{\mu_i(y|w)}} \tag{8}$$

Henceforth, we assume that the primitives μ, λ_H and λ_L are such that

$$\frac{\lambda_L}{\lambda_H} < \sqrt{\frac{\mu_i(x|w)}{\mu_i(y|w)}} < \frac{\lambda_L + 1}{\lambda_H - 1} \tag{9}$$

for every w and every i, such that $q^*(w)$ is an interior solution for every w. As θ_L and ε approach zero, this condition becomes less restrictive.

3.2 Incentive Compatibility

In order to implement the optimal policy, the platform needs to know advertisers' types. Now suppose that the platform is unable to directly verify this information. Therefore, it relies on the advertisers' self-reports - which we interpret as requests to target a specific consumer-preference group. The reports are submitted ex-ante, once and for all - that is, we do not allow for dynamic reporting, in line with our restriction to stationary environments.

A policy (q, F) is incentive-compatible (IC) if no single advertiser has an incentive to misreport its type, given that every other advertiser reports truthfully. When a single advertiser of type x pretends to be y, it changes its probability of display from $q_i(x|w)/m$ to $q_i(y|w)/(m+1)$ for each consumer i. Hence, the transition probability from D_x to S_x changes to

$$\theta_H[q_i(x|w) + \frac{q_i(y|w)}{m+1}] + \theta_L q_i(y|w)(\frac{m}{m+1})$$

since a type x consumer will transact with probability θ_H if the displayed ad is either by one of the truthful x advertisers or by the single deviating advertiser, and with probability θ_L if the displayed ad is by one of the truthful y advertisers. Similarly, the transition probability from D_y to S_y changes to

$$\theta_H[q_i(y|w)(\frac{m}{m+1})] + \theta_L[q_i(x|w) + \frac{q_i(y|w)}{m+1}]$$

Consequently, the invariant probability that consumer i is in state D_x is

$$\tilde{\rho}_i(x|w) = \frac{\mu_i(x|w)\varepsilon}{(\frac{m}{m+1})(\theta_H - \theta_L)q_i(x|w) + \theta_H(\frac{1}{m+1}) + \theta_L(\frac{m}{m+1}) + \varepsilon}$$

and the invariant probability that he is in state D_y is

$$\tilde{\rho}_i(y|w) = \frac{\mu_i(y|w)\varepsilon}{(\frac{m}{m+1})(\theta_H - \theta_L)q_i(y|w) + \theta_L + \varepsilon}$$

In a similar manner, we can derive the invariant probabilities when a single y advertiser deviates. Note that $\tilde{\rho}_i \to \rho_i$ as $m \to \infty$.

It follows that an x advertiser weakly prefers to report its type if and only if

$$\begin{aligned} & \sum_{i \in N} \sum_{w \in W} \mu(w) \frac{q_i(x|w)}{m} [\theta_H \rho_i(x|w) + \theta_L \rho_i(y|w)] - F_x \\ \geq & \sum_{i \in N} \sum_{w \in W} \mu(w) \frac{q_i(y|w)}{m+1} [\theta_H \tilde{\rho}_i(x|w) + \theta_L \tilde{\rho}_i(y|w)] - F_y \end{aligned}$$

We refer to this inequality as the $IC(x, y)$ constraint. The IC constraint of a y advertiser, referred to as $IC(y, x)$, is similarly defined.[4]

We wish to derive conditions under which the optimal policy (q^*, F^*) is IC in the $m \to \infty$ limit. When it is, we say that the optimal policy 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

$$\begin{aligned} \sum_{i \in N} \sum_{w \in W} \mu(w) q_i(y|w)[\rho_i(x|w) - \rho_i(y|w)] & \leq 0 \\ \sum_{i \in N} \sum_{w \in W} \mu(w) q_i(x|w)[\rho_i(y|w) - \rho_i(x|w)] & \leq 0 \end{aligned} \tag{10}$$

Plugging the solution for q^* from the previous sub-section, we can obtain a necessary

[4] In principle, the platform could design a more general mechanism that exploits its knowledge of m, such that all advertisers are severely punished if the number of x reports is not m. This would make truthful reporting trivially incentive-compatible. However, this device is very artificial and unreasonably relies on exact knowledge of m.

and sufficient condition for implementability of the optimal policy in the $m \to \infty$ limit. 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 learned the type of a particular consumer i. Then, we could update our beliefs regarding the overall structure of the social network. The conditional distributions $(\mu_i(w|t_i)_{w\in W}$, $t_i = x, y$, 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 optimal policy. Define

$$S_i \equiv \sum_{w\in W} \sqrt{\mu_i(w|x)\mu_i(w|y)} \tag{11}$$

In the Statistics and Machine Learning literatures, S_i is known as the *Bhattacharyya Coefficient* that characterizes the distributions $\mu_i(\cdot|x)$ and $\mu_i(\cdot|y)$.[5] From a geometric point of view, this is an appropriate similarity measure because S_i is the direction cosine between two unit vectors in $\mathbb{R}^{|W|}$, $(\sqrt{\mu_i(w|x)})_{w\in W}$ and $(\sqrt{\mu_i(w|y)})_{w\in W}$. 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 $\mu_i(\cdot|x)$ and $\mu_i(\cdot|y)$, according to which the realized network w serves as a signal that indicates the types of individual consumers. Indeed, the stochastic matrix $(\mu_i(\cdot|t))_{t\in\{x,y\}}$ 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 $(\mu_i(\cdot|t))_{t\in\{x,y\}}$.*

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.

Example 1 revisited

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

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,

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

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

As an aside, we note that the Bhattacharyya Coefficient is useful in understanding how the platform's payoff from the optimal policy (q^*,F^*) depends on network's informativeness. If we plug the solution (7) into the advertisers' total surplus (5), we obtain:

$$n\varepsilon\left[\frac{\bar{\theta}+\frac{1}{2}\varepsilon}{\bar{\theta}+\varepsilon}-\varepsilon\frac{\sqrt{\pi(1-\pi)}}{\bar{\theta}+\varepsilon}S\right]$$

where $\bar{\theta}=\frac{1}{2}(\theta_H+\theta_L)$. This expression for the platform's "first-best" payoff decreases with the Bhattacharyya Coefficient S when π is held fixed (however, we will see below that S typically varies with π). The intuition is that a more informative network facilitates effective targeting and therefore increases the average number of transactions per period.

Condition for implementing the optimal policy

The next result employs the Bhattacharyya Coefficient to derive a simple necessary and sufficient condition for implementability of the optimal policy. The following result - as well as all subsequent ones - focuses on implementability *in the $m\to\infty$ limit.*

Proposition 1 *Suppose that $q^*(w)$ is an interior solution for every w. Then, (q^*,F^*) is implementable if and only if*

$$S\leq\left(\frac{\lambda_H}{\lambda_L+\lambda_H}\right)\sqrt{\frac{1-\pi}{\pi}}+\left(\frac{\lambda_L}{\lambda_L+\lambda_H}\right)\sqrt{\frac{\pi}{1-\pi}}\qquad(12)$$

Thus, implementability of the optimal policy depends on two factors: the "popularity ratio" $\pi/(1-\pi)$ between the two product types, and the extent to which the social network is *informative* of consumer types (as captured by the Bhattacharyya Coefficient of μ). Recall the analogy to "welfare theorems" described in the Introduction. Proposition 1 can be viewed as a characterization of environments in which competitive markets can sustain an efficient allocation of advertisers over consumers' scarce attention.

To get an intuition for the result, consider the $\lambda_L/\lambda_H \to 0$ limit, where a consumer rarely buys a product from a poor-match advertiser, and where repeat purchases are rare, such that condition (12) simplifies into

$$S\sqrt{\frac{\pi}{1-\pi}} \leq 1 \tag{13}$$

In the $\lambda_L/\lambda_H \to 0$ limit, the optimal display probability $q_i^*(t|w)$ is proportional to the *square root* of $\mu_i(t|w)$. By comparison, the fee paid by a firm that submits the report t is proportional to $\mu(t)$. Thus, although a product with high $\mu_i(t|w)$ 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 type 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 $\mu_i(t|w)$ 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).

The probability ε of exiting the satiation state and the match-quality parameters θ_L and θ_H contribute to the coefficient $\lambda_L/(\lambda_L+\lambda_H)$ that features in condition (12). Note that $\lambda_L/(\lambda_L+\lambda_H) \in (0, \frac{1}{2})$ and that it increases with λ_L/λ_H, implying that it *increases* with ε and θ_L but *decreases* with θ_H. Because $\pi \geq \frac{1}{2}$, an increase in $\lambda_L/(\lambda_L+\lambda_H)$ leads to an increase in the R.H.S of (12), and therefore makes the condition easier to satisfy. The following result summarizes the comparative statics of the necessary condition w.r.t λ_L/λ_H.

Proposition 2 *If the optimal policy is not implementable for a given λ_L/λ_H, then it is not implementable under $\lambda'_L/\lambda'_H < \lambda_L/\lambda_H$.*

In particular, as consumers become *more attentive* to ads or the propensity for repeat purchases *declines*, the necessary and sufficient condition for implementing the optimal policy becomes *harder* to meet.

Example 2: Impossibility with two-node homophily
Consider the simple case in which there are only two consumers who form a link if and only if they are of the same type. There are only two possible networks in this case (the two nodes are either linked or not) and $S = 2\sqrt{\pi(1-\pi)}$. Since $\pi > \frac{1}{2}$, we have that $S > \sqrt{(1-\pi)/\pi}$, hence there exists $\varepsilon^* > 0$ such that for all $\varepsilon < \varepsilon^*$, condition (12) is violated. It follows that when the frequency of repeat purchases is sufficiently low, the optimal policy is not incentive compatible for this simple network-generating process. □

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 $k = 2$, the type distribution σ is represented by π, 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 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 known as *community detection* (see Mossel et al. (2012), 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 optimal policy if the network-formation process obeys an SBM.

When analyzing social networks, a natural question that arises is what causes 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 analysis of the SBM model will occasionally make use of two properties of the Bhattacharyya Coefficient. For the following two remarks, suppose that we can represent w as a pair, $w = (g_1, g_2)$, where $g_k \in G_k$.

Remark 2 *Suppose that* $\mu(g_1, g_2|t) \equiv \mu(g_1|t)\mu(g_2|t)$ *- i.e.,* g_1 *is independent of* g_2 *conditional on* t*, according to* μ*. For every* $k = 1, 2$*, define*

$$S_k = \sum_{g_k \in G_k} \sqrt{\mu(g_k|x)\mu(g_k|y)}$$

Then, $S = S_1 \cdot S_2$*.*

Remark 3 *Suppose that* $\mu(g_1, g_2|t) \equiv \mu(g_1)\mu(g_2|g_1, t)$ *- i.e.,* g_1 *is independent of* t *according to* μ*. For every* g_1*, define*

$$S(g_1) = \sum_{g_2 \in G_2} \sqrt{\mu(g_2|g_1, x)\mu(g_2|g_1, y)}$$

Then,

$$S = \sum_{g_1} \mu(g_1)S(g_1)$$

Remark 2 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 signals' coefficients. Remark 3 says that when one signal is an independent chance move that merely determines the distribution of the other signal conditional on the consumer's type, the Bhattacharyya Coefficient is a weighted average of the latter signals' coefficients. The properties follow immediately from the coefficient's definition, and therefore the proof is omitted.

Our first main 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 optimal policy is not implementable when λ_L/λ_H is small.

Proposition 3 *Fix $n \geq 2$ and a generic P. There exist $\pi^*, \pi^{**} \in (\frac{1}{2}, 1)$ with the property that for every $\pi \in (\pi^*, 1) \cup (\frac{1}{2}, \pi^{**})$ and every sufficiently small λ_L/λ_H, the optimal policy 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, conditional on learning $(t_j)_{j \neq i}$, the link status between any $j, h \neq i$ has no informational content regarding t_i (follows from the assumption that the SBM is known and from Remark 1). 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, Remark 2 implies that 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.

For any δ, we can find π^* sufficiently close to one such that $\sqrt{(1-\pi^*)/\pi^*} = \delta^2 < 1$. For any $\pi > \pi^*$, let $\sqrt{(1-\pi)/\pi} = \hat{\delta}^2$ where $\hat{\delta} < \delta$, and choose the ratio λ_L/λ_H to be sufficiently close to zero such that the R.H.S of (12) is arbitrarily close to $\hat{\delta}^2$, and therefore below δ, thus violating (12).

(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 $p_{xy} = 0$. 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 the two nodes' 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 w.r.t 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 $n-1$ conditionally independent signals: each signal generates a link with probability π $(1-\pi)$ conditional on the original node's type being x (y). By Remark 2, 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 optimal policy:

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

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

$$2^{n-1}\pi^{\frac{n}{2}}(1-\pi)^{\frac{n}{2}-1} \leq (\frac{1}{\lambda_L/\lambda_H+1}) + (\frac{\lambda_L/\lambda_H}{\lambda_L/\lambda_H+1})(\frac{\pi}{1-\pi}) \tag{14}$$

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 $\tau^{**}(\pi)$ such that condition (14) is violated for every $\lambda_L/\lambda_H < \tau^{**}(\pi)$. To show this, it suffices to construct π^{**} and $\tau^{**}(\pi)$ such that for every $\pi \in (\frac{1}{2}, \pi^{**})$ and $\lambda_L/\lambda_H < \tau^{**}(\pi)$, the derivative w.r.t. π of the L.H.S of (14) is strictly higher than the corresponding derivative of the R.H.S.

The derivative of the L.H.S of (14) w.r.t π is equal to

$$2^{n-1}\pi^{\frac{n}{2}-1}(1-\pi)^{\frac{n}{2}-2}[\frac{n}{2} - \pi(n-1)] \tag{15}$$

which is positive if $\frac{1}{2} < \pi < \frac{n}{2(n-1)}$. Since the expression (15) equals 2 when $\pi = \frac{1}{2}$, it is strictly above one when π is sufficiently close to $\frac{1}{2}$.

The derivative of the R.H.S. of (14) w.r.t. π is equal to

$$(\frac{\lambda_L/\lambda_H}{\lambda_L/\lambda_H+1}) \cdot \frac{1}{(1-\pi)^2}$$

which, for all $\pi \in (\frac{1}{2}, 1)$, is positive and increasing in π and λ_L/λ_H. Given $\pi \in (\frac{1}{2}, \pi^{**})$, let $\tau^{**}(\pi)$ be the solution to the equation

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

Hence, for any $\pi \in (\frac{1}{2}, \pi^{**})$ and any $\lambda_L/\lambda_H < \tau^{**}(\pi)$, the derivative w.r.t π of the L.H.S of (14) is strictly higher than the corresponding derivative of the R.H.S. ■

Thus, an intermediate popularity gap is necessary for implementing the optimal policy under the SBM (in the low λ_L/λ_H regime). 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 with high α and low β, 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 difficult 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 optimal policy 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, the optimal policy is not implementable under any SBM with $n = 2$ when λ_L/λ_H is sufficiently small. This observation brings us to the role of n. The following result provides a sufficient condition for implementability of the optimal policy.

Proposition 4 *For any generic (π, P), there exists n^* such that the optimal policy 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, Remark 2 implies that the Bhattacharyya Coefficient that corresponds to this signal is

$$[\sqrt{\eta_x \eta_y} + \sqrt{(1-\eta_x)(1-\eta_y)}]^{n-1} \tag{16}$$

Now, observe that this signal is weakly less informative than learning the entire network w. Therefore, S is weakly below the expression (16). It follows that the following inequality is a sufficient condition for the implementability of the optimal policy:

$$[\sqrt{\eta_x \eta_y} + \sqrt{(1-\eta_x)(1-\eta_y)}]^{n-1} \leq (\frac{1}{\lambda_L/\lambda_H + 1})\sqrt{\frac{1-\pi}{\pi}} + (\frac{\lambda_L/\lambda_H}{\lambda_L/\lambda_H + 1})\sqrt{\frac{\pi}{1-\pi}} \tag{17}$$

For generic (π, P), $\eta_x \neq \eta_y$, such that $\sqrt{\eta_x \eta_y} + \sqrt{(1-\eta_x)(1-\eta_y)} < 1$. In addition, for any $(\pi, \lambda_L, \lambda_H)$ that satisfy (9), the R.H.S. of (17) 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 optimal policy. 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. As $n \to \infty$, the network becomes arbitrarily informative, such that S becomes arbitrarily close to zero, and the condition for implementability of the optimal policy is satisfied.

To get a quantitative sense of Proposition 4, consider the following table, which provides values of n^* for various specifications of the homophily case:

π	α	β	λ_L/λ_H	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	0.2	231
0.75	0.1	0.02	0.2	72

This table illustrates the forces that affect implementability of the optimal policy: 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_L/λ_H.

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 sparser 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 optimal policy will be easier to implement.

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 $P_{\log}$ 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 optimal policy. Specifically, we borrow existing necessary and sufficient conditions for (asymptotic) *exact recovery* of two asymmetric "communities". By exact recovery, we mean that for a given large 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. This implies that S is close to zero and therefore the condition for implementability of the optimal policy holds.

Proposition 5 *In the $n \to \infty$ limit, the optimal policy is implementable if*

$$\pi(a-b)^2 + (1-\pi)(c-b)^2 \geq 2 \tag{18}$$

Proof. By definition,

$$S = \frac{1}{\sqrt{\pi(1-\pi)}} \sum_{w \in W} \mu(w) \sqrt{\mu(x|w)\mu(y|w)}$$

Exact recovery means that the probability (measured according to μ) of realizations w for which $\mu(x|w)$ or $\mu(y|w)$ are arbitrarily close to zero is arbitrarily high. Therefore, exact recovery is ensured if $S \to 0$ when $n \to \infty$.

Let $n \to \infty$. Given the preceding paragraph, we only need 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 (18). ■

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 $n \to \infty$ limit, the optimal policy is implementable in the homophily case if*

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

while in the extroversion/introversion case, the optimal policy 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 optimal policy for a large network is that the homophily or extroversion/introversion effects are strong enough.

5 Informed Advertisers

So far, we assumed that advertisers are entirely uninformed of the realization of 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 optimal policy?

To address the first question, suppose that before the advertiser submits its report to the platform, it receives a signal s regarding the realization of w. The signal is independent of $(t_1, ..., t_n)$ conditional on w. Let r be the joint distribution over networks w and signals s, such that $r(s|w)$ is the probability that an advertiser receives the signal s conditional on the realized network being w. Conversely, let $r(w|s)$ 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. 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 optimal policy is unaffected if it also requires advertisers 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 (10) are exactly the same, except that the term $\mu(w)$ is replaced with $r(w|s)$. We require advertisers' IR constraint to bind *ex-ante* - i.e., on average across signal realizations.

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

$$\sum_{w \in W} r(w|s) \sum_{i \in N} q_i(t|w)[\rho_i(t|w) - \rho_i(t'|w)] \leq 0 \tag{19}$$

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

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

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

Proposition 6 *(i) If the optimal policy is implementable under r, then it is implementable under any r' that is less informative than r. (ii) Suppose $\mu(x|w^*) \neq \frac{1}{2}$ or $\mu(x|w_*) \neq \frac{1}{2}$. Then, if advertisers are fully informed of w (i.e., $r(w|w) = 1$ for every w), the optimal policy is not implementable when λ_L/λ_H is sufficiently small.*

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 optimal policy when λ_L/λ_H 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 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 gets information by sampling a random subset of no more than d nodes (out of the total of n nodes in the network), and learning the subgraph of w over these d nodes. 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 d nodes. If*

$$S(n-d) \leq [\frac{\lambda_H}{\lambda_L+\lambda_H}\sqrt{\frac{\pi}{1-\pi}} + \frac{\lambda_L}{\lambda_L+\lambda_H}\sqrt{\frac{1-\pi}{\pi}}] - [\frac{d}{n-d}\cdot\frac{\sqrt{2}-1}{2\sqrt{\pi(1-\pi)}}] \tag{20}$$

then the optimal policy is implementable.

Proof. Suppose an advertiser learns the subgraph of 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 optimal policy is that for every signal g_1,

$$\begin{aligned}
& \sum_{g_2,h}\mu(g_2,h|g_1)\sum_{i\in N}\sqrt{\mu_i(x|g_1,g_2,h)\mu_i(y|g_1,g_2,h)} \qquad (21)\\
\leq\ & \sum_{g_2,h}\mu(g_2,h|g_1)\sum_{i\in N}\left[(\frac{\lambda_H}{\lambda_H+\lambda_L})\mu_i(y|g_1,g_2,h)+(\frac{\lambda_L}{\lambda_H+\lambda_L})\mu_i(x|g_1,g_2,h)\right]
\end{aligned}$$

and

$$\begin{aligned}
& \sum_{g_2,h}\mu(g_2,h|g_1)\sum_{i\in N}\sqrt{\mu_i(x|g_1,g_2,h)\mu_i(y|g_1,g_2,h)} \qquad (22)\\
\leq\ & \sum_{g_2,h}\mu(g_2,h|g_1)\sum_{i\in N}\left[(\frac{\lambda_H}{\lambda_H+\lambda_L})\mu_i(x|g_1,g_2,h)+(\frac{\lambda_L}{\lambda_H+\lambda_L})\mu_i(y|g_1,g_2,h)\right]
\end{aligned}$$

(These expressions are easily derived from the inequality (24) given at the beginning of the proof of Proposition 1 - see the Appendix.)

Since g_1 and g_2 are independent, we can write $\mu(g_2, h|g_1) = \mu(g_2)\mu(h|g_1, g_2)$. Also, observe that $\mu_i(x|g_1, g_2) = \sum_h \mu(h|g_1, g_2)\mu_i(x|g_1, g_2, h)$. Applying the Cauchy-Schwartz inequality, we obtain

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

It follows that inequalities (21)-(22) are implied by the following, simpler inequalities:

$$\begin{aligned}
\sum_{i\in N}\left[\sum_{g_2}\mu(g_2)\sqrt{\mu_i(x|g_1, g_2)\mu_i(y|g_1, g_2)} - (\frac{\lambda_H}{\lambda_H + \lambda_L})\mu_i(y|g_1) - (\frac{\lambda_L}{\lambda_H + \lambda_L})\mu_i(x|g_1)\right] &\leq 0 \\
\sum_{i\in N}\left[\sum_{g_2}\mu(g_2)\sqrt{\mu_i(x|g_1, g_2)\mu_i(y|g_1, g_2)} - (\frac{\lambda_H}{\lambda_H + \lambda_L})\mu_i(x|g_1) - (\frac{\lambda_L}{\lambda_H + \lambda_L})\mu_i(y|g_1)\right] &\leq 0
\end{aligned}$$

Consider the top inequality (it will be easy to see that if it holds, the bottom inequality holds as well). 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 $\mu_i(x|g_1, g_2) = \mu_i(x|g_1)$. Similarly, for every $i \in N_2$ we can write $\mu_i(x|g_1, g_2) = \mu_i(x|g_2)$ and $\mu_i(x|g_1) = \mu_i(x) = \pi$. It follows that the inequality can be rewritten as

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

The top sum can be simplified into

$$n_2 S(n_2)\sqrt{\pi(1-\pi)} - n_2(\frac{\lambda_H}{\lambda_H+\lambda_L})\pi - n_2(\frac{\lambda_L}{\lambda_H+\lambda_L})(1-\pi)$$

while the bottom sum can be grouped together as

$$\begin{aligned}
&\sum_{i\in N_1}\left[\sqrt{\mu_i(x|g_1)\mu_i(y|g_1)}-(\frac{\lambda_H}{\lambda_H+\lambda_L})\mu_i(x|g_1)-(\frac{\lambda_L}{\lambda_H+\lambda_L})\mu_i(y|g_1)\right] \\
\leq\ & n_1\cdot\max_{\chi\in\{0,1\}}\max_{\varphi\in[0,1]}\left[\sqrt{\varphi(1-\varphi)}-\chi\varphi-(1-\chi)(1-\varphi)\right] \\
=\ & n_1\cdot\frac{\sqrt{2}-1}{2}
\end{aligned}$$

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

$$n_2\left[S(n_2)\sqrt{\pi(1-\pi)}-(\frac{\lambda_H}{\lambda_H+\lambda_L})\pi-(\frac{\lambda_L}{\lambda_H+\lambda_L})(1-\pi)\right]+n_1\frac{\sqrt{2}-1}{2}\leq 0 \tag{23}$$

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

Note that the term in the first bracket on the R.H.S. of (20) is precisely the R.H.S. of (12), the necessary and sufficient condition for implementing the optimal policy, while the term in the second bracket is some positive constant that increases in d. Thus, condition (20) says that the optimal policy is implementable even when advertisers observe a subgraph of the network - as long as the informativeness of the subgraph they do *not* observe is sufficiently above the threshold for implementability. This reflects the fact that it is harder to implement the optimal policy when advertisers have some knowledge of the network.

When π and the connectivity matrix are fixed, inequality (20) is stated entirely in terms of d and n. We can therefore express $S(n-d)$ as a function of d, and use the upper bounds on $S(k)$ that we derived in Section 4 to get a closed-form upper bound on d, such that the optimal policy is implementable for any value of d below that bound. Finally, the comparative statics w.r.t d are consistent with our previous results. When d increases, the R.H.S of (20) clearly goes down, whereas $S(n-d)$ goes up because a smaller network is a less informative signal. Thus, a larger d 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 observed information about consumers (demographic characteristics, actions they took within the confines of the social network). The following is an example of how this extension may change our 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 of w over N_1 and N_2, denoted w_1 and w_2, are 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 $(a_1, ..., a_n)$. By Remark 2, 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, by Remark 3:

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

We can now plug this expression for S into necessary or sufficient conditions for implementability of the optimal policy.

Many preference/product types

Throughout this paper, we assumed that there are only two preference/product types, x and y. Now suppose that there are $K > 2$ types, denoted $x_1, ..., x_K$. Suppose that the high-quality match probability θ_H applies whenever firms' and consumers' types coincide, and that the low-quality match probability θ_L applies in any other case. Consider the case in which the optimal display rule is interior: $q_i^*(x_k|w) > 0$ 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 optimal policy is that for every pair of types x_k and x_{k^*},

$$S(k, k^*) \leq (\frac{\lambda_H}{\lambda_H + \lambda_L})\sqrt{\frac{\mu(x_k)}{\mu(x_{k^*})}} + (\frac{\lambda_L}{\lambda_H + \lambda_L})\sqrt{\frac{\mu(x_{k^*})}{\mu(x_k)}}$$

where $\mu(x_k)$ is the ex-ante probability that a consumer is of type x_k, and $S(k, k^*)$ is the Bhattacharyya Coefficient of $(\mu(w|x_k))_{w \in W}$ and $(\mu(w|x_{k^*}))_{w \in W}$.

7 Conclusion

We presented a modeling framework that sheds light on forces that shape incentive-compatible targeted 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. In particular, our necessary and sufficient condition for implementing the first-best makes the role of the network's informativeness transparent, via its use of the Bhattacharyya Coefficient. This in turn facilitates understanding of the role of various features of the network generating process. The results also demonstrate the non-trivial effects of consumers' attentiveness to advertising and their propensity for repeat purchases. Finally, the analysis highlights - and to some extent quantifies - the importance of keeping advertisers uninformed of the structure of the social network.

Our analysis focused on the platform's "first-best" policy and its implementability, thus leaving open the question of characterizing the "second-best" policy when the first-best is not implementable. This is a challenging question that may require new analytical techniques, because it is not clear which participation and incentive constraints will be binding, and because the second-best solution may break the independence across network realizations that characterizes the first-best solution and its implementability condition. Addressing these challenges is therefore left for future research.

Perhaps the most intriguing aspect of our model is its 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 social networks.

Appendix: Proofs

Proposition 1

From (9), it follows that (7) characterizes the optimal display policy. Plugging this expression for $q_i(t|w)$ into the $IC(x,y)$ constraint (10) yields the following inequality

$$\begin{aligned} & \sum_{i\in N}\sum_{w\in W}\mu(w)\cdot\sqrt{\mu_i(x|w)\mu_i(y|w)} \\ \leq \quad & \sum_{i\in N}\sum_{w\in W}\mu(w)\cdot[(\frac{\lambda_H}{\lambda_L+\lambda_H})\mu_i(y|w)+(\frac{\lambda_L}{\lambda_L+\lambda_H})\mu_i(x|w)] \end{aligned} \tag{24}$$

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

$$\sum_{i\in N}\sum_{w\in W}\sqrt{\mu_i(x,w)\mu_i(y,w)}\leq n(\frac{\lambda_H}{\lambda_L+\lambda_H})(1-\pi)+n(\frac{\lambda_L}{\lambda_L+\lambda_H})\pi \tag{25}$$

Because $\mu_i(x,w)=\pi\mu_i(w|x)$ and $\mu_i(y,w)=(1-\pi)\mu_i(w|y)$, we can express (25) as the following inequality,

$$\sum_{i\in N}\sum_{w\in W}\sqrt{\mu_i(w|x)\mu_i(w|y)}\leq n(\frac{\lambda_H}{\lambda_L+\lambda_H})\sqrt{\frac{1-\pi}{\pi}}+n(\frac{\lambda_L}{\lambda_L+\lambda_H})\sqrt{\frac{\pi}{1-\pi}}$$

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

$$S\leq\frac{\lambda_H}{\lambda_L+\lambda_H}\sqrt{\frac{1-\pi}{\pi}}+\frac{\lambda_L}{\lambda_L+\lambda_H}\sqrt{\frac{\pi}{1-\pi}} \tag{26}$$

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

$$S\leq\frac{\lambda_H}{\lambda_L+\lambda_H}\sqrt{\frac{\pi}{1-\pi}}+\frac{\lambda_L}{\lambda_L+\lambda_H}\sqrt{\frac{1-\pi}{\pi}}$$

By assumption, $\pi\geq\frac{1}{2}$. And since $\lambda_H>\lambda_L$, the only inequality that matters is (26), which is precisely the condition (12).

Proposition 6

(i) The proof is entirely rudimentary and standard. Nevertheless, we give it for completeness. By assumption, inequality (19) holds for every s. Using the definition of Blackwell informativeness, we can rewrite $r'(w|s)$ as

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

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

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

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

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

It follows that for every s, $r'(w|s)$ is some convex combination of $(r(w'|s))_{w'}$. Therefore, given that under r, (19) 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 optimal policy are that for every w,

$$\begin{aligned}
\sum_{i\in N} \sqrt{\mu_i(x|w)\mu_i(y|w)} &\leq \sum_{i\in N} \left[\frac{\lambda_H}{\lambda_L + \lambda_H} \mu_i(y|w) + \frac{\lambda_L}{\lambda_L + \lambda_H} \mu_i(x|w) \right] \\
\sum_{i\in N} \sqrt{\mu_i(x|w)\mu_i(y|w)} &\leq \sum_{i\in N} \left[\frac{\lambda_H}{\lambda_L + \lambda_H} \mu_i(x|w) + \frac{\lambda_L}{\lambda_L + \lambda_H} \mu_i(y|w) \right]
\end{aligned}$$

Let $w \in \{w_*, w^*\}$. Then, w is a symmetric signal - i.e., $\mu_i(x|w)$ 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

$$1 \leq (\frac{1}{\lambda_L/\lambda_H+1})\sqrt{\frac{\mu(y|w)}{\mu(x|w)}}+(\frac{\lambda_L/\lambda_H}{\lambda_L/\lambda_H+1})\sqrt{\frac{\mu(y|w)}{\mu(x|w)}} \quad (27)$$

$$1 \leq (\frac{1}{\lambda_L/\lambda_H+1})\sqrt{\frac{\mu(x|w)}{\mu(y|w)}}+(\frac{\lambda_L/\lambda_H}{\lambda_L/\lambda_H+1})\sqrt{\frac{\mu(x|w)}{\mu(y|w)}} \quad (28)$$

Because $\mu(x|w^*) \neq \frac{1}{2}$ or $\mu(x|w_*) \neq \frac{1}{2}$, either $\mu(x|w) > \mu(y|w)$ or $\mu(x|w) > \mu(y|w)$. Assume w.l.o.g that $\mu(x|w^*) > \mu(y|w^*)$. Since inequality (28) is violated for $\lambda_L/\lambda_H = 0$, there exists $\lambda_L^*/\lambda_H^* > 0$ such that this inequality would also be violated for all $\lambda_L/\lambda_H < \lambda_L^*/\lambda_H^*$.

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] Bergemann, Dirk and Alessandro Bonatti (2015): "Selling cookies," *American Economic Journal: Microeconomics* 7(3), 259-294.

[6] 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.

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

[8] Eliaz, Kfir and Ran Spiegler (2015): "Search design and broad matching," *American Economic Review* 105, 563-586.

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

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

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

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

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

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

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