

# Incentive-Compatible Advertising on Non-Retail Platforms\*

Kfir Eliaz and Ran Spiegler<sup>†</sup>

September 4, 2018

## Abstract

Non-retail platforms (e.g., online radio stations, social media) enable users to engage in various non-commercial activities, while at the same time generating information about users that helps advertisers improve their targeting. We study novel incentive issues that arise when the platform tries to maximize its advertising revenues. In our model, the platform’s policy consists of a personalized, stationary ad-display rule and an advertising fee (which the platform charges 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 platform obtains more precise information about users’ preferences. We apply our result to various examples of non-retail platforms, including content platforms and social networks. Our analysis of the latter turns out to be related to the “community detection” problem in Network Science.

---

\*This is a substantial revision of a paper formerly circulated (and promoted on YouTube by a movie trailer) under the title “Incentive-Compatible Advertising on a Social Network”. 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, numerous seminar audiences and the referees of this journal, for helpful comments. We are especially grateful to the editor, David Myatt, for fruitful suggestions.

<sup>†</sup>Eliaz: School of Economics, Tel-Aviv University and Economics Dept., Aarhus University, 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

Recent years have seen a proliferation of online institutions that can be described as “non-retail platforms”. Users of these platforms access them on a regular basis, in order to engage in activities such as reading texts, listening to music, exchanging messages, cultivating social links, etc. In particular, when they access the platform, it is *not* for the purpose of buying from advertisers. If a user buys from a firm as a result of being exposed to an ad posted on the platform, the transaction takes place off it; it will have no effect on his activity on the platform, and it is quite likely that the platform does not even monitor whether the transaction has taken place. However, the transaction may temporarily depress the user’s demand for similar products, thus diminishing the effectiveness of advertising them.

Of course, non-retail platforms are at least as old as the village message board. What is special about the modern online version is that users’ activity on the platform leaves a massive trail of information that may be correlated with their consumption tastes in various areas. As a result, the platform can help advertisers achieve better targeting, which in turn helps the platform increase its advertising revenues. Here are a few examples of what we have in mind.

*Online radio* stations like Pandora collect information about users’ musical tastes (in this respect, they differ from traditional radio), and can use that to target ads for unrelated products. For instance, whether a user likes Country music may be correlated with his politics and lifestyle preferences. But of course, he does not access Pandora for the purpose of being informed about political candidates or buying vegan food.

*E-mail* services may use the content of personal e-mails to target users. If a user’s e-mails start featuring numerous references to babies, he may experience increased exposure to diaper ads on his e-mail account, although buying diapers is obviously not the user’s primary objective when checking his e-mail.

*Messaging platforms* such as Whatsapp or Snapchat may be unable or unwilling to use the content that users generate, for technical or legal reasons. However, the structure of the social network among users may provide information about their types. For instance, if users exhibit *homophily* - i.e., they associate with like-minded individuals - then a large cluster in the network indicates that its members are likely to have similar tastes.

*Content sharing* platforms such as Reddit are message boards that publish user-generated content, and may monitor the content that users produce or consume.

Many platforms exhibit combinations of these features. For instance, social media platforms like Instagram or Twitter can make use of the network structure of their users as well as the content that they generate. While not all of these real-life examples of non-retail platforms currently make use of this form of targeted advertising, the potential to do so is inherent in them.<sup>1</sup>

In this paper we study novel incentive issues that arise in advertising on non-retail platforms. The source of the potential incentive problem is that advertisers have private information regarding the consumer-preference types they would like to target. The platform relies on their targeting requests to allocate display ads to individual users, utilizing its own private information about users. When advertising fees vary with the targeting request, it becomes a *strategic* decision, which involves trading off the likelihood of a transaction against the fee. Indeed, real-life ad-tech intermediaries help advertisers cope with such trade-offs by searching for the target audience that gives the “best bang for the buck”. This may involve diverting the client’s ad to a less-than-ideal audience because it may be significantly cheaper.<sup>2</sup>

Because users’ ad-generated (offline) purchases affect their willingness to make a subsequent purchases without having any visible effect on their activity at the non-retail platform, it may be desirable for the platform to *diversify* the type of ads it shows to an individual user. Even if a Country-music fan is relatively unlikely to be interested in vegan food, exposing him to such ads every once in a while may increase the long-run expected number of transactions generated by such a user. The paper’s basic insight is that the diversification motive creates an incentive for advertisers to misrepresent their ideal targeting. Our aim is to understand the conditions in which this incentive problem prevents the platform from attaining its first-best.

In our model, there is a group of consumers who are constantly present on some non-retail platform. Each consumer comes 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. The platform obtains a noisy aggregate signal about the profile of consumers’ types. It then enables advertisers (a.k.a firms) to post personalized display ads. Each firm is

---

<sup>1</sup>Search engines are an example of hybrid platforms with both retail and non-retail features. Consumers use search engines to specifically look for a product or service to buy, but they also use them to find information unrelated to any transaction. We do not address such platforms in this paper.

<sup>2</sup>AdEspresso.com is a company that offers to help small businesses launch advertising campaigns on social media. On their website they write, “The audience you choose will directly affect how much you’re paying...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.

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. This is the firm’s private information (in the main version of our model, firms receive no additional information about consumers). Ex-ante, each firm communicates to the platform the type of consumers it wishes to target. Thus, ads are classified into “types” according to the targeting request that accompanies them. Advertiser-platform communication of this kind exists in reality. For instance, Pandora offers ad targeting based (among other things) on users’ music preferences or listening habits.

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 posterior belief regarding his type (derived from its signal) - such that the ad he is exposed to at any period is drawn independently according to this mixture. Ads are like “billboards” and transactions occur 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 platform 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, and calculated according to consumers’ personal Markov processes - and to extract it by means of advertising fees. Because the platform is uncertain about 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. As mentioned above, this turns out to generate a motive for advertisers to *strategize* their targeting request.

Our first observation 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 platform’s posterior probability 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: reduced ad exposure is more than compensated for by the reduced fee.

Under what conditions on the environment’s primitives can the platform design an incentive-compatible policy (consisting of a display rule and an advertising-fee schedule) 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 regard the non-retail platform as a market institution for allocating consumers’ limited attention (the scarce resource in this environment) to advertisers. The full-surplus-extraction requirement is a zero-profit condition that captures competitive behavior among advertisers. Our question then becomes: *Can an efficient allocation of platform users’ attention to advertisers be supported by a competitive market?* We do *not* study “second-best” policies when the first-best is not implementable: this is a challenging problem that requires different analytical techniques and belongs to a different paper.

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 - otherwise, our condition is merely sufficient). The condition is an inequality that incorporates two quantities: (i) on the L.H.S., a measure of the informativeness of the platform’s signal; and (ii) on the R.H.S, a simple expression that involves the characteristics of the consumer population.

The chief merit of this characterization is that it isolates the platform-specific details and summarizes them by the L.H.S’s informativeness measure. The inequality’s R.H.S summarizes the consumers’ features that are independent of the platform. Therefore, examining various types of platforms is reduced to studying their induced informativeness measure. Another virtue of the inequality is that it makes comparative statics quite transparent. The inequality is *easier to satisfy* when the signal 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*. However, it is important to emphasize that in all the applications we examine in this paper, the two sides of the inequality are *interdependent* - i.e., the signal’s informativeness varies with the consumer-type distribution. The reason is that the platform brings together multiple consumers, such that the platform’s signal about an individual consumer depends on the types of other platform users he interacts with.

We apply our basic result to three specifications that capture different non-retail platforms. Our first two applications are extremely simple “warm-up” exercises that deal with e-mail and content platforms. The third and main application examines a social network, where the only information that is available to the platform is the

network structure, and the probability of a link between two users is a function of their types. We show that in this context, the first-best is not implementable if the consumer type distribution is either too asymmetric or too uniform. We then ask whether a larger network makes it easier for the platform to implement its objective. Following the Network Science literature on *community detection*, we assume that users’ propensity to form links decreases with network size, such that the expected degree of an individual node grows only *logarithmically* in  $n$ . Applying a recent result on the community-detection problem (Abbe and Sandon (2015)), we obtain a sufficient condition for the implementability of the platform’s objective for large  $n$  in terms of parameters of the network-generating process. 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. Thus, unlike non-retail platforms, the search platform’s *sole* function is to match users with advertisers. Eliaz and Spiegler (2015) borrowed the Bhattacharyya Coefficient (a measure of similarity between probability distributions) from the statistics and machine learning literature, and demonstrated its use in representing IC constraints. The present paper further demonstrates the power of this tool in a different context, and with new technical challenges that arise from the applications (e.g., the community detection problem in 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 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 social network. Recent papers in this strand include Galeotti and Goyal (2012) and Campbell (2015) (see Bloch (2015) for a survey). In these papers, consumers are targeted according to their network *location*, whereas in

the present paper targeting is based on the platform’s information regarding their *preference types*.

Bergemann and Bonatti (2015) study a different aspect of using information about consumers for advertising purposes. In their model, a single data provider offers firms information about the potential value of matches with various consumers, where the information is obtained from consumers’ online activities.

## 2 A Model

We begin with a single-consumer environment. As will become clear, this is without loss of generality for the basic results in Section 3, and the extension to multiple consumers will be immediate.

Let  $T = \{x, y\}$  be a set of possible consumer types. We consider a stationary environment in which a consumer of type  $t \in T$  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 uninterested in consumption. The consumer’s transition between states obeys the following mechanical rule. He 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  $\varepsilon$ . 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 the consumer buys a product conditional on being exposed to its ad while in his demand state is  $\theta_H$  when the product’s type coincides with his own type, and  $\theta_L$  when it does not, where  $\theta_L < \theta_H$ . The parameters  $\theta_L, \theta_H$  also reflect the consumer’s 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 of the types (our analysis will focus on the case of large  $m$ ). Each advertiser can costlessly supply any amount of its product. If the consumer acquires a product from an advertiser, the advertiser earns a fixed payoff of 1. In particular, our model completely abstracts from product prices.<sup>3</sup>

The consumer has constant, uninterrupted access to a non-retail platform that receives a signal  $w$  from a finite set  $W$ , where  $|W| > 1$ . The signal is received ex-ante,

---

<sup>3</sup>For examples of models that study the effect of platform features on firms’ pricing, see Eliaz and Spiegler (2011), Candogan et al. (2012) and Fainmesser and Galeotti (2015).

once and for all, before the above stochastic process that describes consumer behavior begins. The platform does not receive any information about the types of individual advertisers. Advertisers receive no information about the consumer's type - we relax this assumption in Section 5.2. Let  $\mu \in \Delta(T \times W)$  be a prior joint distribution over the consumer's type and the platform's signal. Then,  $\mu(t | w)$  is the platform's posterior probability that the consumer is of type  $t$  conditional on receiving the signal  $w$ . We often use the abbreviated notation  $\phi_w = \mu(t = x | w)$ . Let  $\pi = \sum_w \mu(x, w)$  be the prior probability that the consumer is of type  $x$ , where  $\pi \geq \frac{1}{2}$ , without loss of generality.

How does the platform match advertisers and consumers? At every time period, 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 the platform cannot monitor them. In particular, there is no notion of “clicking” on display ads.

The display of ads is governed by a stationary rule that the platform commits to ex-ante. (In an environment with multiple consumers, the ad would be personalized - i.e., there will be a distinct ad display process for each consumer.) Formally,  $q_w^t$  is the probability that at any time period, the platform displays an advertiser of type  $t$  conditional on the signal  $w$  (of course,  $q_w^x + q_w^y = 1$ ). Conditional on displaying an ad of type  $t$ , each of the type- $t$  advertisers is drawn with equal probability. Hence, the probability that a particular advertiser of this type is displayed is  $q_w^t/m$ . We refer to  $q_w = (q_w^x, q_w^y)$  as the platform's *display rule* for  $w$ . Let  $F_t$  be the per-period fee the platform charges 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 signal  $w$  and a display rule  $q$ , the transition probabilities between the mental states of a consumer of type  $t$  are given by the following matrix:

$$\begin{array}{cc}
 & \begin{array}{c} D_t \\ S_t \end{array} \\
 \begin{array}{c} D_t \\ S_t \end{array} & \begin{array}{cc}
 & \begin{array}{c} D_t \\ S_t \end{array} \\
 1 - [\theta_H q_w^t + \theta_L (1 - q_w^t)] & \theta_H q_w^t + \theta_L (1 - q_w^t) \\
 \varepsilon & 1 - \varepsilon
 \end{array}
 \end{array} \tag{1}$$

Hence, given  $w$  and  $q$ , the joint invariant probability that the consumer is of type  $t$



and in state  $D_t$  is

$$\rho_w^t \equiv \frac{\mu(t | w)\varepsilon}{\theta_H q_w^t + \theta_L(1 - q_w^t) + \varepsilon} \quad (2)$$

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. For simplicity, we imagine that the possible vectors 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.

*The firms' private information.* One could argue that the platform need not rely on advertisers' targeting requests - in principle, it could examine each advertiser's product and figure out the quality of its match with each consumer type. However, this ad-classification task is costly, and the platform can avoid the cost by decentralizing the 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 packages that fit "outdoorsy" tourists are quite blurred. In these cases, market studies are likely to reveal information that the platform lacks. It is implausible for the platform to replicate such studies in the myriad industries it interacts with.

*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  $w$ , 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 the consumer pays attention to ads and whether he transacts

with firms (e.g., he may be listening to Pandora while jogging). Therefore, it cannot learn anything about consumers beyond the signal  $w$ . Even so, stationary display rules (which constitute the endogenous aspect) carry a 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. The optimal display rule would involve switching between  $x$  and  $y$  ads, but the precise transition rule - and therefore, the long-run number of transactions - would be extremely hard to characterize. We believe that the qualitative insights of our stationary model would not change, as long as we assume that advertisers do not know the initial state of the Markov process - they would still treat the allocation of ad slots at any given period as a random variable, albeit one whose distribution is difficult to calculate. The stationarity assumption is thus a simplifying approximation that enables us to tractably capture the platform's motive to diversify its ad types over time. However, the quality of this approximation will vary with  $\theta_L, \theta_H, \varepsilon$ , and this means that our comparative statics with respect to these parameters should be taken with a grain of salt.

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.

### 3 Basic Results

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

#### 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 display rule that maximizes advertisers' surplus. To calculate the gross expected per-period payoff (without taking into account any payment to the platform) for an advertiser of type  $t$ , we multiply the invariant prob-

ability that the consumer is in his demand state by the probability that he transacts conditional on being in this state. It follows that the expected number of transactions per period with advertisers of type  $x$  is

$$U_x(q) \equiv \sum_{w \in W} \mu(w) q_w^x [\theta_H \rho_w^x + \theta_L \rho_w^y] \quad (3)$$

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

$$U_y(q) \equiv \sum_{w \in W} \mu(w) q_w^y [\theta_H \rho_w^y + \theta_L \rho_w^x] \quad (4)$$

Let  $\bar{q}$  be the display rule that maximizes the sum

$$U_x(q) + U_y(q) \quad (5)$$

We refer to  $\bar{q}_w$  as the optimal display rule for the aggregate signal realization  $w$ . The optimal fee that the platform charges advertisers of type  $t$ , denoted  $\bar{F}_t$ , fully extracts the maximal surplus of these advertisers - i.e.,  $\bar{F}_t = U_t(\bar{q})$ . When  $\bar{q}_w$  is interior for all  $w$ , first-order conditions imply

$$\frac{\rho_w^x}{\rho_w^y} = \sqrt{\frac{\mu(x | w)}{\mu(y | w)}} = \sqrt{\frac{\phi_w}{1 - \phi_w}} \quad (6)$$

Substituting the R.H.S. of (2) for  $\rho_w^t$  allows us to solve explicitly for  $\bar{q}_w$ :

$$\bar{q}_w^x = \frac{(\theta_H + \varepsilon)\sqrt{\phi_w} - (\theta_L + \varepsilon)\sqrt{1 - \phi_w}}{(\theta_H - \theta_L)(\sqrt{\phi_w} + \sqrt{1 - \phi_w})} \quad (7)$$

Denote

$$\lambda = \frac{\theta_H + \varepsilon}{\theta_H + \theta_L + 2\varepsilon} \in \left(\frac{1}{2}, 1\right) \quad (8)$$

As  $\lambda \rightarrow 1$ ,

$$\bar{q}_w^x \rightarrow \frac{\sqrt{\phi_w}}{\sqrt{\phi_w} + \sqrt{1 - \phi_w}} \quad (9)$$

Henceforth, we assume that the primitives are such that the solution is interior for all  $w$ . Equivalently:

$$\frac{1 - \lambda}{\lambda} < \sqrt{\frac{\phi_w}{1 - \phi_w}} < \frac{\lambda}{1 - \lambda} \quad (10)$$

for every  $w$ . This condition becomes more easier to satisfy when  $\lambda$  increases (reflecting

low propensity for repeat purchases and a large effect of product match on consumers' attentiveness) and when the platform's posterior beliefs are relatively uniform. The latter property will hold when the platform's prior is not too asymmetric and its signal is not too informative.

The latter observation highlights the role of consumer satiation in our model. Suppose that consumers experienced no satiation at all, such that their behavior would be described by a single-state process in which they demand their favorite product at every period. Then, the platform's optimal display rule would be a corner solution: if  $\mu(t | w) > \frac{1}{2}$ , it would display type- $t$  ads with probability one at every period. This case is partially approached in our two-state model when  $\varepsilon = 1$ , such that the consumer's satiation lasts exactly one period. In this case, condition (10) holds for the smallest set of other primitives  $\mu, \theta_L, \theta_H$ . Thus, our two-state Markov process is the simplest formalization of the aspect of consumer behavior (namely, satiation) that gives rise to interior optimal display rules, which in turn create the incentive-compatibility problem that will be the subject of the next sub-section.

#### *Signal informativeness and first-best payoffs*

Let us now derive an expression for the platform's payoff induced by the optimal policy - under the assumption that the optimal display rule is always interior. Before doing so, let us introduce the following quantity, which is a function of the distribution  $\mu$ :

$$I = \sum_{w \in W} \mu(w) \sqrt{\mu(x | w) \mu(x | w)} = \mathbb{E} \left[ \sqrt{\phi_w (1 - \phi_w)} \right]$$

The quantity  $I \in [0, \frac{1}{2}]$  is the expectation of a concave function of  $\phi_w$ , the platform's posterior probability that the consumer's type is  $x$ . Recall the property of Bayesian posteriors:

$$\pi = \sum_{w \in W} \mu(w) \phi_w$$

Consider a change in  $\mu$  that keeps  $\pi$  fixed and makes the signal  $w$  more informative in Blackwell's sense. This is equivalent to introducing a mean-preserving spread to the distribution over  $\phi_w$ . Jensen's inequality then implies that  $I$  decreases. Therefore, it can be viewed as a measure of the *informativeness of the platform's signal*.<sup>4</sup>

Now, plugging the solution (7) into the advertisers' total surplus (5), we obtain

$$\varepsilon \left[ 1 - \frac{\varepsilon(1 + 2I)}{\theta_H + \theta_L + 2\varepsilon} \right] \tag{11}$$

---

<sup>4</sup>We thank the editor, David Myatt, for suggesting this measure.

This expression for the platform’s “first-best” payoff decreases with  $I$ . The intuition is that a more informative signal facilitates effective targeting and therefore increases the average number of transactions per period.

To summarize this sub-section, the optimal display policy  $\bar{q}$  is given by (7) whenever the primitives satisfy the condition for interior solutions (10), and the induced first-best payoff is given by (11).

### 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 advertisers’ self-reports - which we interpret as requests to target a specific 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.<sup>5</sup> When a single advertiser of type  $x$  pretends to be  $y$ , it changes its probability of display from  $q_i(x | w)/m$  to  $q_i(y | w)/(m + 1)$ . Hence, the transition probability from  $D_x$  to  $S_x$  changes to

$$\theta_H(q_w^x + \frac{q_w^y}{m + 1}) + \theta_L \frac{mq_w^y}{m + 1}$$

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

$$\theta_H \frac{mq_w^y}{m + 1} + \theta_L(q_w^x + \frac{q_w^y}{m + 1})$$

Consequently, the invariant probability that consumer  $i$  is of type  $x$  and in state  $D_x$  is

$$\tilde{\rho}_w^x = \frac{\mu(x | w)\varepsilon}{(\frac{m}{m+1})(\theta_H - \theta_L)q_w^x + \theta_H(\frac{1}{m+1}) + \theta_L(\frac{m}{m+1}) + \varepsilon}$$

---

<sup>5</sup>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 the invariant probability that he is in state  $D_y$  is

$$\tilde{\rho}_w^y = \frac{\mu(y | w)\varepsilon}{\binom{m}{m+1}(\theta_H - \theta_L)q_w^y + \theta_L + \varepsilon}$$

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

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

$$\sum_{w \in W} \mu(w) \frac{q_w^x}{m} [\theta_H \rho_w^x + \theta_L \rho_w^y] - F_x \geq \sum_{w \in W} \mu(w) \frac{q_w^y}{m+1} [\theta_H \tilde{\rho}_w^x + \theta_L \tilde{\rho}_w^y] - F_y$$

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. When these two constraints are satisfied, we say that the optimal policy is implementable.

We focus on the  $m \rightarrow \infty$  limit. Equivalently, we could allow  $m$  to be finite and assume that advertisers are “price takers”, in the sense that they do not take into account the effect of their own behavior on the transition probabilities in the consumer’s Markov process. Because  $\bar{F}$  fully extracts advertisers’ surplus, the L.H.S of  $IC(x, y)$  and  $IC(y, x)$  is zero. In the  $m \rightarrow \infty$  limit, the inequalities thus reduce to

$$\begin{aligned} \sum_{w \in W} \mu(w) q_w^y [\rho_w^x - \rho_w^y] &\leq 0 \\ \sum_{w \in W} \mu(w) q_w^x [\rho_w^y - \rho_w^x] &\leq 0 \end{aligned} \tag{12}$$

Plugging the solution for  $\bar{q}$  from the previous sub-section, we obtain a simple necessary and sufficient condition for implementability of the optimal policy *in the  $m \rightarrow \infty$  limit*.

**Proposition 1** *Suppose that  $\bar{q}_w$  is interior for every  $w$ . Then,  $(\bar{q}, \bar{F})$  is implementable in the  $m \rightarrow \infty$  limit if and only if*

$$I \leq (1 - \lambda)\pi + \lambda(1 - \pi) \tag{13}$$

Recalling 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 consumers’ scarce attention to advertisers. This result highlights two factors. As the platform’s signal becomes more informative (measured by  $I$ ) and as the consumer type distribution (given by  $\pi$ ) becomes more

symmetric, condition (13) becomes easier to satisfy (when  $\pi = \frac{1}{2}$ , the condition necessarily holds because  $I \leq \frac{1}{2}$  by definition). It should be emphasized, however, that in all the applications we will examine in Section 4, these two factors are *not* independent - i.e., a change in  $\pi$  will also affect the informativeness of the platform’s signal, an interdependency that will have non-trivial effects. Still, distinguishing between these two factors is helpful for understanding Proposition 1.

Consider the  $\lambda \rightarrow 1$  limit, where a consumer rarely buys a product from a poor-match advertiser and where the propensity for repeat purchases is low. Condition (13) simplifies into

$$I \leq 1 - \pi \tag{14}$$

The optimal display probability  $\bar{q}_w^t$  in the  $\lambda \rightarrow 1$  limit is proportional to the *square root* of  $\mu(t | w)$ . By comparison, the fee paid by a firm that submits the report  $t$  is proportional to  $\mu(t)$ . Thus, although a product with high  $\mu(t | w)$  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.

Fixing  $I$ , when the consumer type distribution becomes more uneven (such that  $1 - \pi$  goes down), the gap between the fees paid by firms of different types widens, and this exacerbates the misreporting incentive. As the signal becomes more informative, the values of  $\mu(t | w)$  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 signal 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.

Let us now move away from the  $\lambda \rightarrow 1$  regime. The probability  $\varepsilon$  of exiting the satiation state and the match-quality parameters  $\theta_L$  and  $\theta_H$  determine the weights of  $\pi$  and  $1 - \pi$  in the R.H.S of (13). Because  $\lambda \in (\frac{1}{2}, 1)$  and  $\pi \geq \frac{1}{2}$ , a decrease in  $\lambda$  relaxes the inequality (13) and makes the optimal policy easier to implement. More specifically, as consumers become *more attentive* to ads (in the sense that  $\theta_L$  and  $\theta_H$  increase by the same factor), and as the propensity for repeat purchases *declines*, the condition for implementing the optimal policy becomes *harder* to meet.

## 4 Applications

So far, our analysis focused on the case of a single, isolated consumer. Yet, non-retail platforms typically operate by bringing together multiple consumers, whose *interaction* reveals information about their individual characteristics. Friendship patterns in a social network or the content of an e-mail exchange are cases in point. In order to capture this feature of non-retail platforms in our applications, we now assume that there are *multiple consumers* and that the platform obtains an *aggregate* signal about their individual types. This requires a slight adaptation of our formalism.

Specifically, the set of consumers is  $\{1, \dots, n\}$  and  $t_i \in T_i = \{x, y\}$  denotes the type of consumer  $i$ . Denote  $T = T_1 \times \dots \times T_n$ . The prior distribution  $\mu$  is defined over  $T \times W$ . Consumers' types are *i.i.d.*, where  $\pi \geq \frac{1}{2}$  continues to denote the prior probability that  $t_i = x$ . The conditional signal distribution  $\mu(w | t)$  treats consumers symmetrically: for every permutation  $f : T \rightarrow T$  there exists a permutation  $g : W \rightarrow W$  such that  $\mu(w | t) = \mu(g(w) | f(t))$ . As a result, the distribution over the platform's posterior belief regarding the type of any consumer is the same for all consumers.

Thanks to the latter observation, the analysis in Sections 2 and 3 is easily adapted to this multi-consumer setting. First, we replace  $\mu(t | w)$  with  $\mu(t_i | w)$ , which represents the probability that consumer  $i$ 's type is  $t_i$  conditional on the aggregate signal  $w$ . Accordingly, we replace the notation  $\phi_w$  with  $\phi_{i,w}$ , which represents the probability that consumer  $i$ 's type is  $x$  conditional on the signal  $w$ . Second, we replace the notation  $q_w^t$  with the notation  $q_{i,w}^t$ , which represents the probability that an ad of type  $t$  is displayed at any period to consumer  $i$  given the aggregate signal  $w$ . Once these adaptations are made, the optimal display rule is given by (7) for all consumers; expression (11) describes the platform's first-best payoff *per consumer*; and most importantly for the purposes of this section, the implementability condition (13) is completely unchanged, once the definition of  $I$  replaces  $\mu(t | w)$  with  $\mu(t_i | w)$ .

### *The Bhattacharyya Coefficient*

Another minor adaptation that is useful for the applications concerns the measure of the informativeness of the platform's signal. While  $I$  is a transparent measure, the fact that it is defined in terms of the distribution over the platform's posterior belief makes it cumbersome in applications, because it requires us to compute this distribution. Fortunately, we can slightly rewrite this measure such that it is defined entirely in terms of the conditional distribution  $\mu(w | t_i)$ , which is easier to derive from primitives.



Define

$$S \equiv \sum_{w \in W} \sqrt{\mu(w | t_i = x)\mu(w | t_i = y)} \quad (15)$$

The symmetry of  $\mu$  with respect to consumers' labels implies that  $S$  is the same for all consumers  $i$ .

In the Statistics and Machine Learning literatures,  $S$  is known as the *Bhattacharyya Coefficient* that characterizes the distributions  $\mu(\cdot | x)$  and  $\mu(\cdot | y)$ .<sup>6</sup> It is a measure of *similarity* between these two conditional distributions. From a geometric point of view, this is an appropriate similarity measure because  $S$  is the direction cosine between two unit vectors in  $\mathbb{R}^{|W|}$ ,  $(\sqrt{\mu(w | x)})_{w \in W}$  and  $(\sqrt{\mu(w | y)})_{w \in W}$ . The value of  $S$  increases as the angle between these two vectors shrinks;  $S = 1$  if the two vectors coincide; and  $S = 0$  if they are orthogonal.

More importantly,  $S$  is a measure of the informativeness of the platform's signal. The stochastic matrix  $(\mu(\cdot | t_i))_{t_i \in \{x, y\}}$  can be viewed as an information system in Blackwell's sense. Eliaz and Spiegler (2015) established that  $S$  decreases with the Blackwell informativeness of this information system. Indeed, elementary algebra establishes an immediate connection between  $S$  and our earlier measure of signal informativeness:

$$I = S\sqrt{\pi(1 - \pi)}$$

This enables us to restate the condition for the implementability of the platform's optimal policy:

$$S \leq (1 - \lambda)\sqrt{\frac{\pi}{1 - \pi}} + \lambda\sqrt{\frac{1 - \pi}{\pi}} \quad (16)$$

This is the condition that will serve us in the remainder of this section. In Section 5 we examine extensions of the model, which demonstrate additional advantages of using the measure  $S$ .

The rest of this section is structured as follows. In Sections 4.1 and 4.2 we examine simple "warm-up" examples, before turning to a weightier application in Section 4.3. Throughout the section, we restrict attention to the large  $\lambda$  regime - i.e., the case of low propensity for repeat purchases and low consumer attention to low-match products. This simplifies expressions without changing the substance of our analysis. We also take it for granted that the entire analysis is conducted for the  $m \rightarrow \infty$  limit.

---

<sup>6</sup>See Basu, Shioya and Park (2011) and Theodoris and Koutroumbas (2008). A related concept is the *Hellinger distance* between distributions, given by  $H^2(x, y) = 1 - \sqrt{S(x, y)}$ .

## 4.1 Targeting Based on E-Mail Content

Consider two users of an e-mail service. A user of type  $x$  is expecting a baby or has a newborn, whereas a user of type  $y$  does not. This may affect their preferences over a wide range of products. For example, type  $x$  is more likely to be interested in home entertainment and diapers, whereas type  $y$  is more likely to be interested in concerts or alcohol. The platform that operates the e-mail service monitors users' *sent mail* folder. The aggregate signal  $w$  is a pair  $w = (w_1, w_2)$ , where  $w_i = x$  indicates that the users' sent e-mails contain references to babies and  $w_i = y$  indicates that they do not. The two users are friends who exchange messages with each other, and the platform's algorithm cannot tell whether a user refers to his own baby or his friend's. Moreover, assume that a user of type  $x$  always sends e-mails that mention his baby, while a user of type  $y$  mentions babies with positive probability  $\xi \in (0, 1)$  *only* when his friend has one. The conditional signal distribution  $\mu((w_1, w_2) \mid (t_1, t_2))$  is given by the following table:

$(t_1, t_2) \backslash (w_1, w_2)$	$x, x$	$x, y$	$y, x$	$y, y$
$x, x$	1	0	0	0
$x, y$	$\xi$	$1 - \xi$	0	0
$y, x$	$\xi$	0	$1 - \xi$	0
$y, y$	0	0	0	1

Let us now derive a condition for implementability of the first-best in this example. Consider the distribution over aggregate signals conditional on user 1's type. First, the signals  $(y, y)$  and  $(y, x)$  are impossible when  $t_1 = x$  because by assumption, such a user type always sends an e-mail with reference to babies. Second, the signal  $(x, y)$  is impossible when  $t_1 = y$ , because this babyless type can only refer to babies if his friend has a baby, which would then imply  $w_2 = x$ , a contradiction. It follows that the only signal realization that contributes a non-zero term to the Bhattacharyya Coefficient is  $w = (x, x)$ . Therefore,

$$S = \sqrt{\mu((x, x) \mid t_1 = x)\mu((x, x) \mid t_1 = y)} = \sqrt{(\pi + (1 - \pi)\xi) \cdot \pi\xi}$$

This expression demonstrates how a change in the consumer type distribution (given by  $\pi$ ) can also affect the informativeness measure  $S$ . The reason is that the platform's information about an individual user depends on his platform-mediated interaction with the other user, which in turn depends on his type.

Condition (16) in the  $\lambda \rightarrow 1$  limit is thus

$$\sqrt{(\pi + (1 - \pi)\xi) \cdot \pi\xi} \leq \sqrt{\frac{1 - \pi}{\pi}}$$

which simplifies into the following result.

**Proposition 2** *The first-best is implementable in the  $\lambda \rightarrow 1$  limit if and only if*

$$\pi^3\xi(1 - \xi) + \pi^2\xi + \pi \leq 1$$

We can see that if  $\pi$  is sufficiently close to  $\frac{1}{2}$ , the condition holds for all  $\xi$ . However, for every sufficiently high  $\pi$ , there exists  $\xi^*(\pi)$  such that the condition fails for all  $\xi > \xi^*(\pi)$ . Thus, a more symmetric distribution of consumer types is unambiguously better for implementability of the first-best. As we will later see, this is not always the case. Also, as babyless users become less likely to mention babies in their e-mails, the platform's signal gets more informative, and the condition for implementability becomes easier to satisfy.

## 4.2 Content Platforms

In a prevalent type of non-retail platforms, users upload and consume content. In the case of online newspapers, users' role is almost entirely limited to consumption. Other platforms treat users as both producers and consumers of content. In some cases (e.g., Reddit), the platform acts as a communal message board. Others (e.g. Pinterest) contain elements of a social network. One common feature of these content platforms is that the content that a user consumes reflects not only his personal taste but also the *availability* of various types of content (and therefore, indirectly, other users' types). In particular, a user may fail to consume his ideal content if it is scarce or not prominent.

In this sub-section we study a very stylized example of advertising on a content platform. There are  $n$  content consumers who face a supply of  $m$  content items. Each item is generated by a producer, whose type is drawn from the same distribution as the consumers, such that a producer of type  $t \in \{x, y\}$  always uploads content of type  $t$ . We ignore the targeting of content producers (because by assumption, they perfectly reveal their type and therefore generate no incentive issues for advertising) and focus entirely on consumers. A consumer always consumes exactly one of the  $m$  available items. A signal about consumer  $i$  specifies the types of each of the  $m$  available content items, and the type of the item that consumer  $i$  consumes. A consumer of type  $t$

consumes type  $t$  content whenever it is available. Hence, the *only* case in which the consumer will *not* consume his ideal type of content is when all  $m$  items are of the other type.

In this set-up, the Bhattacharyya Coefficient is simple to derive. Whenever the signal realization  $w$  is such that both content types are available, each consumer  $i$  consumes his type of content with probability one. In this case, either  $\mu(w | t_i = x) = 0$  or  $\mu(w | t_i = y) = 0$ , such that  $w$  contributes a zero term to the coefficient. It follows that the only signal realizations  $w$  with a positive contribution are those in which only one type of content is available, and in this case consumer choices are completely uninformative of their type - i.e.  $\mu(w | t_i)$  is constant in  $t_i$ . As a result, the contributed term  $\sqrt{\mu(w | t_i = x)\mu(w | t_i = y)}$  is equal to the probability of the realized available content profile. Therefore,  $S = \pi^m + (1 - \pi)^m$ . This immediately implies the following result.

**Proposition 3** *The first-best is implementable in the  $\lambda \rightarrow 1$  limit if and only if*

$$\pi^m + (1 - \pi)^m \leq \sqrt{\frac{1 - \pi}{\pi}}$$

This result captures the intuition that a large supply of content that is drawn from a heterogeneous population of producers increases the chances that consumers' observed content consumption will reveal their tastes, and therefore enhances the informativeness of the platform's signal. Formally, for every  $m > 1$  there is a critical value  $\pi^*(m) \in (\frac{1}{2}, 1)$ , such that the condition for first-best implementability holds if and only if  $\pi \leq \pi^*(m)$ . Note that  $\pi^*(m)$  increases with  $m$  - i.e., a larger supply of content makes it easier to implement the first-best because the uninformative signal realizations become rarer. Since moving  $\pi$  closer to  $\frac{1}{2}$  always improves informativeness, the comparative statics with respect to  $\pi$  are clear-cut.

### 4.3 Social Networks

We now turn to our main application in this paper, where the platform operates a social 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 network structure does not evolve over time. We focus entirely on the informational content of the network structure itself, and ignore other aspects of the users' activity on the network that may generate valuable information for advertisers. This will enable us to establish a theoretical connection with an interesting question in the Network Science literature.

Formally, a social network is a random non-directed graph in which consumers are nodes. The set  $W$  can thus be redefined as the set of all possible networks. From now on, we will refer to elements in  $N$  as consumers or nodes interchangeably. We assume that  $\mu$  obeys a random graph process known as the *stochastic block model* (SBM). An SBM is characterized by a triplet  $(n, \sigma, P)$ , where  $n$  is the number of nodes,  $\sigma$  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  $k = 2$  that fits our model, the type distribution  $\sigma$  is represented by  $\pi$ , 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  $\sigma$  and  $P$  generate a joint distribution  $\mu$  over consumer-type profiles and social networks that treats consumers symmetrically, as assumed at the beginning of this section.

The following are two natural specifications of the  $k = 2$  SBM. Under *homophily*, 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 specification is *extroversion/introversion*: some agents have a greater propensity to form social links than others. This case, too, can be represented with two parameters  $\alpha > \beta$ , such that  $p_x = \alpha^2$ ,  $p_y = \beta^2$  and  $p_{xy} = \alpha\beta$ .<sup>7</sup>

*An example: A three-node network with perfect homophily*

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 the graphs in which exactly two nodes are connected. We can use this observation to calculate  $\mu(w | t_1)$ . For example, the probability that the network is fully connected conditional on  $t_1 = x$  is  $\pi^2$ , 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.

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,

$$S = \sqrt{\mu(w_{ijl} | x)\mu(w_{ijl} | y)} + \sqrt{\mu(w_{jl} | x)\mu(w_{jl} | y)} + 2\sqrt{\mu(w_{ij} | x)\mu(w_{ij} | y)} = 4\pi(1 - \pi)$$

---

<sup>7</sup>Bramouille et al. (2012) develop an alternative model of homophily, in which the number of nodes is random as well.

Therefore, the first-best is implementable in the  $\lambda \rightarrow 1$  limit if and only if

$$4\pi(1 - \pi) \leq \sqrt{\frac{1 - \pi}{\pi}}$$

which simplifies into

$$16\pi^3(1 - \pi) \leq 1$$

This condition is satisfied as long as  $\pi \gtrsim 0.92$ . That is, implementability of the first-best requires a highly *asymmetric* type distribution. Contrast this with our findings in Sub-sections 4.1-4.2, where implementability depended on a relatively *symmetric* type distribution.  $\square$

Our next result relies on the following property of the Bhattacharyya Coefficient.

**Remark 1** *Suppose that for any consumer  $i$ , we can represent  $w$  as a pair,  $w = (g_1, g_2)$ , such that  $\mu(g_1, g_2 | t_i) \equiv \mu(g_1 | t_i)\mu(g_2 | t_i)$ . For every  $k = 1, 2$ , define*

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

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

Remark 1 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. The property follows immediately from the coefficient's definition, and therefore the proof is omitted.

**Proposition 4** *Fix  $n \geq 2$  and a generic  $P$ . There exist  $\pi^*, \pi^{**} \in (\frac{1}{2}, 1)$  with the following property. For every  $\pi \in (\pi^*, 1) \cup (\frac{1}{2}, \pi^{**})$  there exists  $\lambda^*(\pi)$  such that for every  $\lambda > \lambda^*(\pi)$ , 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  $\pi^*$  and  $\pi^{**}$ .

(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  $t_i$  (this 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 1 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  $\delta$ , we can find  $\pi^*$  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  $\lambda$  to be sufficiently close to 1 such that the R.H.S of (16) is arbitrarily close to  $\hat{\delta}^2$ , and therefore below  $\delta$ , thus violating (16).

(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  $\pi$ , 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 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  $n - 1$  conditionally independent signals: each signal generates a link with probability  $\pi(1 - \pi)$  conditional on the original node's type being  $x$  ( $y$ ). By Remark 1, 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:

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

This inequality can be rewritten as follows:

$$2^{n-1}\pi^{\frac{n}{2}}(1 - \pi)^{\frac{n}{2}-1} - \lambda - (1 - \lambda)\left(\frac{\pi}{1 - \pi}\right) \leq 0 \tag{17}$$

The inequality is binding for  $\pi = \frac{1}{2}$ . All we now need to show is that there exists  $\pi^{**} > \frac{1}{2}$  and a function  $\lambda^*(\pi)$  such that for every  $\pi \in (\frac{1}{2}, \pi^{**})$  and every  $\lambda > \lambda^*(\pi)$ , the derivative of (17) with respect to  $\pi$  is strictly positive. Straightforward calculation establishes that this is indeed the case. ■

Thus, a moderately asymmetric consumer type distribution is necessary for implementing the optimal policy under the SBM (in the high  $\lambda$  regime). To get an intuition for this result, recall our discussion of condition (13), which highlighted two interdependent factors that facilitates implementing the first-best: relative symmetry of the consumer type distribution and informativeness of the platform's signal.

The case of a  $\pi$  close to one 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 that is *independent* of  $\pi$ . Therefore, a very asymmetric type distribution overweighs whatever positive effect it may have on the informativeness



factor.<sup>8</sup>

The case of  $\pi$  close to  $\frac{1}{2}$  involves different reasoning. In this case, the network is very uninformative about the nodes' types. For example, in the homophily case with high  $\alpha$  and low  $\beta$ , the network is very likely to consist of two fully connected components, yet these 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  $\pi \rightarrow \frac{1}{2}$  regime, and it is not clear a priori which effect is stronger. However, it turns out that when  $\pi$  is close to  $\frac{1}{2}$ , the uninformativeness factor outweighs the type-distribution symmetry factor.

In the three-node example, we saw that implementing the first-best is impossible for most values of  $\pi$ , even though the SBM was maximally informative given the network size. This raises the question of whether increasing  $n$  would help implementing the first-best. The following result gives a positive answer.

**Proposition 5** *Fix a generic  $(\pi, P)$ . There exists  $n^*$  such that the optimal policy is implementable for all SBMs  $(n, \pi, P)$  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

$$\sqrt{\eta_x \eta_y} + \sqrt{(1 - \eta_x)(1 - \eta_y)} \quad (18)$$

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 1 implies that the Bhattacharyya Coefficient that corresponds to this signal is

$$[\sqrt{\eta_x \eta_y} + \sqrt{(1 - \eta_x)(1 - \eta_y)}]^{n-1} \quad (19)$$

Now, observe that this signal is weakly less informative than learning the entire network  $w$ . Therefore,  $S$  is weakly below the expression (19). It follows that the following

---

<sup>8</sup>This result does not contradict the three-node example, because perfect homophily is not generic. Slight perturbation of the connectivity matrix in the example would lead to non-implementability for sufficiently large  $\pi$ .

inequality is a sufficient condition for the implementability of the optimal policy:

$$[\sqrt{\eta_x \eta_y} + \sqrt{(1 - \eta_x)(1 - \eta_y)}]^{n-1} \leq \lambda \sqrt{\frac{1 - \pi}{\pi}} + (1 - \lambda) \sqrt{\frac{\pi}{1 - \pi}} \quad (20)$$

For generic  $(\pi, P)$ ,  $\eta_x \neq \eta_y$ , such that  $\sqrt{\eta_x \eta_y} + \sqrt{(1 - \eta_x)(1 - \eta_y)} < 1$ . In addition, for any quadruple  $(\pi, \theta_L, \theta_H, \varepsilon)$  that satisfies (10), the R.H.S. of (20) 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. The proof involves a simple “law of large numbers” argument. For illustration, consider 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  $n \rightarrow \infty$ , 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 5, consider the following table, which provides values of  $n^*$  for various specifications of the homophily case:

$\pi$	$\alpha$	$\beta$	$\lambda$	$n^*$
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	0.833	231
0.75	0.1	0.02	0.833	72

This table illustrates the forces that affect implementability of the optimal policy, via their effect on (18) - the Bhattacharyya Coefficient of a signal that indicates whether there is a link between two given nodes (as  $n^*$  is based on this quantity).

Up to now we assumed that the likelihood of forming links does not change as we increase the network size, such that 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 turn to a literature within Network Science known as *community detection* (see Mossel et al. (2012), Abbe and Sandon (2015), and the references therein). The objective in this literature 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. The literature 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 in our model, 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.

Following the practice in the community detection literature, 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

$$p_x = a^2 \frac{\ln(n)}{n} \quad p_{xy} = b^2 \frac{\ln(n)}{n} \quad p_y = c^2 \frac{\ln(n)}{n}$$

where  $a, b, c$  are arbitrary constants. To derive a *sufficient* condition for implementability of the optimal policy, 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 6** *In the  $n \rightarrow \infty$  limit, the optimal policy is implementable whenever*

$$\pi(a - b)^2 + (1 - \pi)(c - b)^2 \geq 2 \tag{21}$$

**Proof.** Recall that

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

for any node  $i$ . Exact recovery means that the probability (measured according to  $\mu$ )

of realizations  $w$  for which  $\mu(t_i = x | w)$  or  $\mu(t_i = y | w)$  are arbitrarily close to zero is arbitrarily high. Therefore, exact recovery is ensured if  $S \rightarrow 0$  when  $n \rightarrow \infty$ .

Let  $n \rightarrow \infty$ . 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

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

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

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

which is equivalent to (21). ■

Note that in the homophily case we have  $a = c$ , whereas the extroversion/introversion case satisfies  $b = \sqrt{ac}$ . Thus, Proposition 6 implies the following.

**Corollary 1** *In the  $n \rightarrow \infty$  limit, the optimal policy is implementable in the homophily case whenever*

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

*whereas in the extroversion/introversion case, the optimal policy is implementable whenever*

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

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 Extensions

In this section we examine two extensions of the basic model.

### 5.1 More than Two Types

Our first extension concerns the number of product/preference types. Throughout this paper, we assume there are only two. Suppose that there are  $K > 2$  types, denoted

$x_1, \dots, x_K$ . Suppose that the high-quality match probability  $\theta_H$  applies whenever firms' and consumers' types coincide, and that the low-quality match probability  $\theta_L$  applies in any other case.

Consider the case in which the optimal display rule is interior - i.e.,  $\bar{q}_w^{x_k} > 0$  for every  $w \in W$  and  $k = 1, \dots, K$ . Then, it is straightforward to show that a necessary and sufficient condition for implementability of the optimal policy is that for every  $k, j = 1, \dots, K$ ,

$$S(k, j) \leq \lambda \sqrt{\frac{\mu(x_k)}{\mu(x_j)}} + (1 - \lambda) \sqrt{\frac{\mu(x_j)}{\mu(x_k)}}$$

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

This exercise also demonstrates the usefulness of the Bhattacharyya Coefficient. In the two-type case, we could use the measure  $I$ , which is defined in terms of the distribution over the platform's posterior belief, because that belief was reduced to a scalar. This is no longer the case when  $K > 2$ , whereas the Bhattacharyya Coefficient continues to serve as a meaningful measure of the extent to which the platform's signal separates a given pair of types.

## 5.2 Partially 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 sub-section is a negative answer to this question. This finding then raises an immediate follow-up question: when advertisers can partially retrieve the platform's information, how much can they learn without destroying the platform's ability to implement the optimal policy?

To address the first question, return to the *single-consumer* case of Sections 2-3, and suppose that before an advertiser submits its report to the platform, it receives a signal  $s$  regarding the realization of  $w$ . The signal is independent of  $t$  conditional on  $w$ . Let  $r$  be the joint distribution over the platform's signal  $w$  and the advertiser's signal  $s$ . We allow the advertisers' signals to be correlated conditional on  $w$ . 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 (12) 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 their signal realizations.

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

$$\sum_{w \in W} r(w | s) q(t | w) [\rho(t | w) - \rho(t' | w)] \leq 0 \quad (22)$$

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.

**Proposition 7** *(i) If the optimal policy is implementable under  $r$ , then it is implementable under any  $r'$  that is less informative than  $r$ . (ii) If advertisers are fully informed of the platform's signal (i.e.,  $r(w | w) = 1$  for every  $w$ ), the optimal policy is not implementable when  $\lambda$  is sufficiently large.*

The reason why releasing information about  $w$  to advertisers cannot help the platform is standard - it means that IC constraints that previously held only on average are now required to hold for all signals. Part (ii) of the result establishes that this monotonicity result is not vacuous: giving advertisers full information about the platform's signal will prevent it from implementing its optimal policy when  $\lambda$  is large.

Suppose that the platform cannot prevent advertisers from learning *part* of its own signal; how much information can it afford to give away? In the remainder of this section, we analyze this question in the context of the *social network* application of Section 4.3. 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,  $k = 1, \dots, n$  (where the connectivity matrix is fixed). Denote this coefficient by  $S(k)$ .

**Proposition 8** *Suppose each advertiser is informed of the subgraph induced by  $w$  over a random subset of at most  $d$  nodes. If*

$$S(n-d) \leq \lambda \sqrt{\frac{\pi}{1-\pi}} + (1-\lambda) \sqrt{\frac{1-\pi}{\pi}} - \frac{d}{n-d} \cdot \frac{\sqrt{2}-1}{2\sqrt{\pi(1-\pi)}} \quad (23)$$

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  $g_1$ ,

$$\begin{aligned} & \sum_{g_2, h} \mu(g_2, h \mid g_1) \sum_{i \in N} \sqrt{\mu(t_i = x \mid g_1, g_2, h) \mu(t_i = y \mid g_1, g_2, h)} \\ & \leq \sum_{g_2, h} \mu(g_2, h \mid g_1) \sum_{i \in N} [\lambda \mu(t_i = y \mid g_1, g_2, h) + (1-\lambda) \mu(t_i = x \mid g_1, g_2, h)] \end{aligned} \quad (24)$$

and

$$\begin{aligned} & \sum_{g_2, h} \mu(g_2, h \mid g_1) \sum_{i \in N} \sqrt{\mu(t_i = x \mid g_1, g_2, h) \mu(t_i = y \mid g_1, g_2, h)} \\ & \leq \sum_{g_2, h} \mu(g_2, h \mid g_1) \sum_{i \in N} [\lambda \mu(t_i = x \mid g_1, g_2, h) + (1-\lambda) \mu(t_i = y \mid g_1, g_2, h)] \end{aligned} \quad (25)$$

(These expressions are easily derived from the inequality (27) given at the beginning of the proof of Proposition 1 - see the Appendix.)

Since  $g_1$  and  $g_2$  are independent, we can write  $\mu(g_2, h \mid g_1) = \mu(g_2) \mu(h \mid g_1, g_2)$ . Also, observe that  $\mu(t_i = x \mid g_1, g_2) = \sum_h \mu(h \mid g_1, g_2) \mu(t_i = x \mid g_1, g_2, h)$ . Applying

the Cauchy-Schwartz inequality, we obtain

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

It follows that inequalities (24)-(25) are implied by the following, simpler inequalities:

$$\begin{aligned} \sum_{i \in N} \left[ \sum_{g_2} \mu(g_2) \sqrt{\mu(t_i = x | g_1, g_2)\mu(t_i = y | g_1, g_2)} - \lambda\mu(t_i = y | g_1) - (1 - \lambda)\mu(t_i = x | g_1) \right] &\leq 0 \\ \sum_{i \in N} \left[ \sum_{g_2} \mu(g_2) \sqrt{\mu(t_i = x | g_1, g_2)\mu(t_i = y | g_1, g_2)} - \lambda\mu(t_i = x | g_1) - (1 - \lambda)\mu(t_i = y | g_1) \right] &\leq 0 \end{aligned}$$

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  $\mu(t_i = x | g_1, g_2) = \mu(t_i = x | g_1)$ . Similarly, for every  $i \in N_2$  we can write  $\mu(t_i = x | g_1, g_2) = \mu(t_i = x | g_2)$  and  $\mu(t_i = x | g_1) = \mu(t_i = x) = \pi$ . It follows that the inequality can be rewritten as

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

The top sum can be simplified into

$$n_2 S(n_2) \sqrt{\pi(1 - \pi)} - n_2 \lambda \pi - n_2 (1 - \lambda)(1 - \pi)$$

while the bottom sum can be grouped together as

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

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:

$$n_2 \left[ S(n_2) \sqrt{\pi(1-\pi)} - \lambda\pi - (1-\lambda)(1-\pi) \right] + n_1 \frac{\sqrt{2}-1}{2} \leq 0 \quad (26)$$

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 (23) is precisely the R.H.S. of (16), 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 (23) 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  $\pi$  and the connectivity matrix are fixed, inequality (23) 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 4.3 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 with respect to  $d$  are consistent with our previous results. When  $d$  increases, the R.H.S of (23) clearly goes down, whereas  $S(n-d)$  goes up because a smaller network is a less informative signal. Thus, a larger  $d$  makes it more difficult to satisfy the sufficient condition.

## Appendix: Omitted Proofs

### Proposition 1

From (10), it follows that (7) characterizes the optimal display policy. Plugging this expression for  $q_w^t$  into the  $IC(x, y)$  constraint (12) yields the following inequality

$$\sum_{w \in W} \mu(w) \cdot \sqrt{\mu(x|w)\mu(y|w)} \leq \sum_{w \in W} \mu(w) \cdot [\lambda\mu(y|w) + (1-\lambda)\mu(x|w)] \quad (27)$$

Note that  $\mu(w)\mu(t|w) = \mu(t, w)$  and  $\sum_{w \in W} \mu(x, w) = \pi$ . The above inequality can thus be rewritten as (16). If we carry out a similar exercise for  $IC(y, x)$ , we obtain the inequality

$$I \leq \lambda\pi + (1-\lambda)(1-\pi)$$

By assumption,  $\lambda, \pi \geq \frac{1}{2}$ . Therefore, the only inequality that matters is (16).

**Proposition 7**

(i) The proof is entirely rudimentary and standard. Nevertheless, we give it for completeness. By assumption, inequality (22) 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  $r$ , (22) 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  $w$ ,

$$\begin{aligned} \sqrt{\mu(x | w) \mu(y | w)} &\leq \lambda \mu(y | w) + (1 - \lambda) \mu(x | w) \\ \sqrt{\mu(x | w) \mu(y | w)} &\leq \lambda \mu(x | w) + (1 - \lambda) \mu(y | w) \end{aligned}$$

Consider a signal realization  $w^*$  for which  $\mu(x | w^*) \neq \frac{1}{2}$  (there must exist such a realization). The above inequalities can be written as

$$1 \leq \lambda \sqrt{\frac{\mu(y | w^*)}{\mu(x | w^*)}} + (1 - \lambda) \sqrt{\frac{\mu(y | w^*)}{\mu(x | w^*)}} \quad (28)$$

$$1 \leq \lambda \sqrt{\frac{\mu(x | w^*)}{\mu(y | w^*)}} + (1 - \lambda) \sqrt{\frac{\mu(x | w^*)}{\mu(y | w^*)}} \quad (29)$$

Because  $\mu(x | w^*) \neq \frac{1}{2}$ , either  $\mu(x | w^*) > \mu(y | w^*)$  or  $\mu(x | w^*) < \mu(y | w^*)$ . Assume

the former, without loss of generality. Since inequality (29) is violated for  $\lambda = 1$ , this inequality would also be violated whenever  $\lambda$  is sufficiently large.

## References

- [1] Abbe, Emmanuel and Colin Sandon (2015): “Community detection in general stochastic block models: fundamental limits and efficient recovery algorithms,” *IEEE 56th Annual Symposium on Foundations of Computer Science*, pp. 670–688.
- [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 versus 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. Bramouille, A. Galeotti and B. Rogers eds.), Oxford University Press.
- [7] Bramouille, Yann, Sergio Currarini, Matthew Jackson, Paolo Pin and Brian Rogers (2012): “Homophily and Long-Run Integration in Social Networks,” *Journal of Economic Theory* 147(5), 1754-1786.
- [8] Campbell, James (2015): “Localized price promotions as a quality signal in a publicly observable network,” *Quantitative Marketing and Economics* 13(1), 27-57.
- [9] Candogan, Ozan, Kostas Bimpikis and Asuman Ozdaglar (2012): “Optimal Pricing in Networks With Externalities,” *Operations Research* 60(4), 883–905.

- [10] Eliaz, Kfir, and Ran Spiegler (2011): “A Simple Model of Search Engine Pricing,” *Economic Journal* 121(556), F329-F339.
- [11] Eliaz, Kfir and Ran Spiegler (2015): “Search design and broad matching,” *American Economic Review* 105, 563-586.
- [12] Fainmesser, Itay and Andrea Galeotti (2015): “Pricing Network Effects,” *Review of Economic Studies* 83(1), 165-198.
- [13] Galeotti, Andrea and Sanjeev Goyal (2012): “Network multipliers: the optimality of targeting neighbors,” *Review of Network Economics* 11(3), 1446-9022.
- [14] Iyer, Ganesh, David Soberman and J. Miguel Villas-Boas (2005): “The targeting of advertising,” *Management Science* 24(3), 461-476.
- [15] Johnson, Justin P. (2013): “Targeted advertising and advertising avoidance,” *Rand Journal of Economics* 44(1), 128-144.
- [16] Mossel, Elchanan, Joe Neeman and Allan Sly (2012): “Stochastic block models and reconstruction,” mimeo.
- [17] Theodoridis, Sergios and Konstantinos Koutroumbas (2008): *Pattern Recognition*. Academic Press.
- [18] Zubcsek, Peter Pal and Miklos Sarvary (2011): “Advertising to a social network,” *Quantitative Marketing and Economics* 9(1), 71-107.