

A Game of Clicks: Economic Incentives to Fight Click Fraud in Ad Networks*

Lemonia Dritsoula
University of California
Santa Cruz, USA
lenia@soe.ucsc.edu

John Musacchio
University of California
Santa Cruz, USA
johnm@soe.ucsc.edu

ABSTRACT

Ad networks use revenue sharing and effective filtering of fraudulent clicks to attract publishers. We develop a simple Hotelling competition-based game-theoretic model to study the effect of competition along these dimensions. We compute the Nash equilibrium strategy for two ad networks that compete for publishers. We then investigate how the preferences of the publishers and the quality of the ad networks affect the market share and the strategies chosen at equilibrium.

Keywords

Duopoly, competition, click fraud, advertising, Hotelling

1. INTRODUCTION

The online advertising market typically involves different classes of players: publishers, ad networks, and advertisers. Advertisers have products to advertise and design the ads. Publishers own websites at which the ads can be placed and receive traffic. Ad networks act as intermediaries between publishers and advertisers: they match the publishers with the different ads, and charge the advertisers for a certain fraction of clicks, the ones the ad network deems valid.

Click fraud (or click spam) has been a serious problem in the online advertising market. By click spam, we define the act of clicking an ad without an interest to see the ad. When such clicks are counted as “valid” by the ad network, the advertiser pays for a useless click, and the publisher is rewarded for generating it. Thus there is an incentive for fraudulent publishers to inflate click numbers. Ad networks have an incentive to identify and filter out fraudulent clicks in order to deliver a more valuable service to advertisers, which are their customers. On the other hand, ad networks do receive revenue for fraudulent clicks, which creates an incentive in the opposite direction to fight fraud less.

Work has been done to gauge the degree of click fraud. For example, Dave et al. [2] have provided a systematic methodology to estimate and measure the click spam in ad networks. Their analysis shows that the click spam is a serious problem, that tends to grow as the mobile advertising market develops.

*This work was supported by NSF grants CNS-090711 and CNS-0953884.

Mungamuru, Weis, and Molina [5] have studied the effects of click fraud in the online advertising market by modeling the incentives of the different actors. The main result of their analysis is that ad networks have a net incentive to fight fraud, despite getting revenue from fraudulent clicks that are billed to the advertisers. They have considered a market of advertisers, publishers, and ad networks, and have concluded that the ad network can gain a market advantage by aggressively combating fraud. Their analysis, though, is a one-step best response analysis and does not result in the computation of Nash equilibria. Our work investigates both the quality of classification algorithms and the revenue share as strategies for the ad networks, who compete for publishers, and we derive the Nash equilibria.

Hotelling has argued in his seminal work [3] that in reality, duopoly is not fragile: a small price advantage by one firm does not capture the whole market. He showed that “some buy from one seller, some from another, in spite of moderate differences of price.” In our paper, we consider a similar “location” model, at which the publishers’ preferences are distributed uniformly on a line between two ad networks.

Kim shows in [4] that in the context of several applications of contemporary importance, the dispersion of consumers relative preferences between competing firms results in softening market competition, and studies how the intensity of competition influences the effects of firms strategies. We also establish a similar result: when the publishers become more heterogeneous in their preferences, the competition in prices becomes less fierce. Researchers use the Hotelling model within models of network platform competitions [6], [9]. Of particular relevance is a model by Njoroge, et al. [7], in which ISPs compete on both price and quality just as we consider ad networks competing in two similar dimensions.

Perlof and Salop [8] have showed that as users’ preferences become more intense, equilibrium price increases. Similarly, changes in the utilities of the publishers by a different multiplicative factor in our model, led to different equilibrium prices.

Comparing to other economic results on competition of identical or differentiated products that take as given the differentiation between the products (horizontal and/or vertical) and examine how price competition takes place under network effects [1], we study the competition between the two ad networks in both price and degree of differentiation.

We are interested in this two-dimensional competition between ad networks — they simultaneously compete on filtering aggressiveness and revenue share given to the publishers. Our model is admittedly simplified, but it still captures



Figure 1: Publishers' preferences are uniformly distributed between AN₁ and AN₂.

those aspects that are interesting when fighting click fraud. To the best of our knowledge, we are the first to investigate the Nash equilibria in games of such a setting. Will one network choose to be tolerant with filtering and compensate by giving a bigger share to publishers, while the other network is more aggressive and gives a smaller share? Will the networks even care fighting click fraud? How are the preferences of the publishers affect the decisions of the ad networks? In which direction will the competition be fiercer?

To address the above questions, our paper is organized as follows. Section 2 describes the underlying economic model. Section 3 provides the Nash equilibrium analysis. Section 4 presents numerical experiments that show how the quality of the ad networks and the distribution on the publishers' preferences affect the revenue sharing strategies of the two ad networks. Finally, we conclude with the main results and insights in Section 5.

2. ECONOMIC MODEL

We consider a one-shot game between two ad networks, called AN₁ and AN₂. The two ad networks compete to receive clicks (display ads) from the publishers. The publishers are uniformly distributed along a line of length 1 between the two ad networks, as shown in Fig. 1. Preferences are driven by anticipated click volume. A publisher could believe that one ad network would be better at placing relevant ads for the type of content the publisher offers and the demographics of the users it serves. In this context, we assume that AN₁ is "preferred" by some publishers and AN₂ is "preferred" by others.

Each ad network, AN_{*i*}, simultaneously decides how aggressively to filter out invalid clicks, and what fraction of the revenue the publishers will get. After both ad networks announce their decisions,

1. Publishers choose between the ad networks, according to their preferences and the revenue they get.
2. Ad networks mark a fraction of these clicks as valid.
3. Advertisers adjust their bids in ad auctions to realize a fixed return on investment – based on the anticipated ratio of truly valid clicks to clicks that are marked valid by the ad network.
4. Advertisers pay for the clicks marked as valid.

Our goal is to compute how aggressive ad networks will be, what fraction of their revenue will be distributed to the publishers at equilibrium, and how the market of publishers will react.

2.1 Ad networks

As in [5], we assume that ad networks can identify fraudulent (invalid) clicks with a receiver operating characteristic (ROC) curve of the form shown in Fig. 2. Ad network *i* is

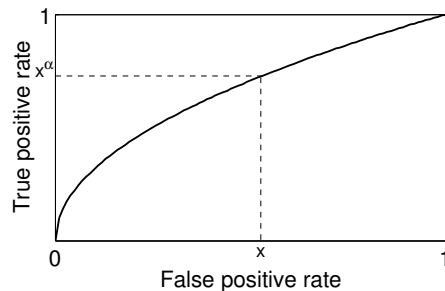


Figure 2: ROC curve: if the ad networks are willing to tolerate a false positive rate of x_i , they can achieve a true positive rate of $x_i^{\alpha_i}$.

endowed with a type α_i that characterizes the ROC curve of his click fraud filtering technology. Each ad network's inherent type is the effectiveness $\alpha_i \in [0, 1]$ of their filtering, while their strategic decisions involve the aggressiveness $x_i \in [0, 1]$ and the revenue share $h_i \in [0, 1]$. We define aggressiveness x_i to be the fraction of valid clicks classified as invalid. An ad network that is more aggressively identifying fraud would choose a higher value x_i . Given x_i and α_i , each ad network marks a fraction $(1 - x_i)$ of valid clicks as valid, and a fraction $(1 - x_i^{\alpha_i})$ of invalid clicks as valid, as shown in Table 1. If $\alpha_1 < \alpha_2$, AN₁ is more effective.

Variable	Classification	Truth	Result
x_i	"invalid"	valid	False positive
$1 - x_i$	"valid"	valid	True negative
$x_i^{\alpha_i}$	"invalid"	invalid	True positive
$1 - x_i^{\alpha_i}$	"valid"	invalid	False negative

Table 1: Ad networks make mistakes when filtering out invalid clicks.

The goal of the ad networks is to maximize their revenue

$$U_i^{\text{AN}}(x_i, h_i) = (1 - h_i) \cdot N_i^r \cdot c,$$

where h_i is the revenue share given to the publishers, c is the price per each ad click, and N_i^r is the number of clicks marked by ad network *i* as valid or "real" (the ones the advertiser is charged for). N_i^r is a function of both the quality α_i of the classification algorithms and the aggressiveness x_i selected by the ad network and is given by

$$N_i^r = (1 - x_i) \cdot r \cdot V_i + (1 - x_i^{\alpha_i}) \cdot (1 - r) \cdot V_i,$$

where r is the fraction of total clicks that are real or valid, and V_i is the volume of the clicks received by each network. The volume V_i depends on how the market of publishers is split.

2.2 Publishers

Depending on the quality of traffic and clicks generated, publishers can either be classified as good or bad. The information is asymmetric: publishers know if they are good or bad, but the ad networks do not. Therefore ad networks need to develop classification algorithms. Also, bad publishers that get discovered as being bad can easily change identities.

We assume that all clicks generated on good publishers' websites are valid, and that all clicks generated on bad pub-

lishers' websites are invalid. The assumption that good publishers have only good clicks is extreme, but it is a convenient way to model the fact that good publishers will have a much larger fraction of good clicks than fraudulent ones. It would be cumbersome to add more parameters, like the fraction of bad clicks for good publishers and the fraction of bad clicks for bad publishers for instance.

We consider that publishers have different preferences for ad networks. Publishers are uniformly distributed along a line $\theta \in [0, 1]$. The point of division between the regions served by the two ad networks (denoted by θ^*) is determined by the condition that at this place the publishers are indifferent between AN₁ and AN₂. Equating the delivered publishers' revenues we have

$$h_1\Phi_1(x_1)[(1-\theta^*)(1-g)+g] = h_2\Phi_2(x_2)[\theta^*(1-g)+g], \quad (1)$$

where $\Phi_i(x_i) = \frac{r \cdot (1 - x_i)}{r \cdot (1 - x_i) + (1 - r) \cdot (1 - x_i^{\alpha_i})}$ is the fraction of charged clicks that are valid, which also depends on the quality of each network α_i .

The parameter $g \in [0, 1]$ is the degree of platform homogeneity. It reflects the importance of the preferences of the publishers with respect to the prices. When g is small ($g = 0$), preferences are more important, while when g is large ($g = 1$), prices have a greater impact on the decision of the publishers. In the subsequent analysis (Section 3), we investigate both extreme cases and highlight how modeling the publishers differently affects the equilibrium strategies of the ad networks. Solving Eq. (1) for θ^* we find

$$\theta^* = \frac{h_1\Phi_1(x_1) - gh_2\Phi_2(x_2)}{(1-g)[h_1\Phi_1(x_1) + h_2\Phi_2(x_2)]}. \quad (2)$$

2.3 Advertisers

As in [5], we assume that the number of advertisers is sufficiently large and covers the number of ad positions on the publishers' websites. This is a realistic assumption, as the advertisers are actually competing to display their ads through auctions. The advertisers adjust their bids to maintain a certain return on investment. It is arguable they would have such a strategy since they would want to invest in online advertising up until the point its return is comparable to that achieved from other forms of advertising. Depending on the quality of the clicks they pay for, they adjust their bids to account for clicks of inferior quality by a factor of

$$\frac{r}{r(1-x_i) + (1-r)(1-x_i^{\alpha_i})}.$$

We avoid dealing with the auctions mechanism details and focus on the implications of the revenue sharing and aggressiveness levels selected by the ad networks. Therefore, we assume that the advertisers' bid is the price per click c , multiplied by the adjustment factor.

3. EQUILIBRIA

Following the previous analysis, the profits of the ad networks are

$$J_1(x_1, h_1) = (1 - h_1) \cdot r \cdot c \cdot V \left(\frac{1}{2} - \frac{\theta^{*2}}{2} \right) \quad (3)$$

Parameter	Definition
α_i	quality of ad network i
h_i	revenue share given to publisher i
x_i	aggressiveness of ad network i
r	fraction of clicks that are valid ("real")
c	price per click
V_i	volume of clicks for ad network i
N_i^r	number of ad network i 's clicks that are valid
g	degree of platform homogeneity
θ^*	point of market segmentation
$\Phi_i(x_i, h_i)$	fraction of charged clicks that are valid

Table 2: List of variables introduced in Section 2

$$J_2(x_2, h_2) = (1 - h_2) \cdot r \cdot c \cdot V \cdot \left(\frac{\theta^*}{2} - \frac{\theta^{*2}}{2} \right), \quad (4)$$

where θ^* is given by Eq. (2).

Each ad network i needs to select $x_i \in [0, 1]$, and $h_i \in [0, 1]$ to maximize the revenues given in Eq. (3) and (4). We first analyze the case when the publishers are heterogeneous: preferences are more important than revenues.

3.1 Heterogeneous publishers ($g = 0$).

In this case, the market share is highly determined by the preferences of the publishers. For example, when a publisher is located near (strongly prefers) AN₁, no matter what revenue share is given, AN₂ will never win over all the market.

LEMMA 1. *The ad networks' payoff functions $J_i(x_i, h_i)$ are concave with respect to x_i, h_i , for $i = 1, 2$.*

PROOF SKETCH. Differentiating twice Eq. (3), we show that $\frac{\partial^2 \theta^*}{\partial x_1^2} > 0$. We thus show the concavity of $J_1(x_1, h_1)$ with respect to x_1 . Similarly, we show the concavity of $J_1(x_1, h_1)$ with respect to h_1 , and of $J_2(x_2, h_2)$ with respect to x_2, h_2 . \square

THEOREM 1. *The levels of aggressiveness chosen by the ad networks at equilibrium are $x_1^* = 1$, and $x_2^* = 1$.*

PROOF. The conditions $\partial^2 J_i(x_i, h_i) / \partial x_i^2 < 0$, sufficient for a maximum of each of the functions $J_i(x_i, h_i)$, $i = 1, 2$, are satisfied (Lemma 1). Solving $\partial J_1 / \partial x_1 = 0$ for x_1 results to $\Phi_1(x_1) \partial \Phi_1(x_1) / \partial x_1$, or $(1 - \alpha_1)x^{\alpha_1} + \alpha_1 x^{\alpha_1 - 1} - 1 = 0$. The unique solution to the previous equation for $x_1 \in [0, 1]$ is $x_1^* = 1$. We can similarly prove that $x_2^* = 1$. \square

3.2 General model for publishers ($g > 0$).

In this case, publishers are distributed between the two ad networks, not only according to their preferences, but according to the revenue share they get as well. We can similarly prove that $x_1^* = 1$ and $x_2^* = 1$.

LEMMA 2. *The payoff function of each ad network is quasi-concave with respect to the revenue share h_i , $i = 1, 2$.*

PROOF. When $g > 0$, there exists a possibility that the ad network with the higher revenue share will win over all the publishers. We have already established that before such a point occurs, the payoff functions will be concave. After this inflection point, the ad network will have already won over all the publishers. Since the competitor is already out of the game, there is no benefit for the winning ad network to increase the revenue share. Thus the payoff function will be decreasing with respect to h_i . Overall, the payoff functions will be quasi-concave. \square

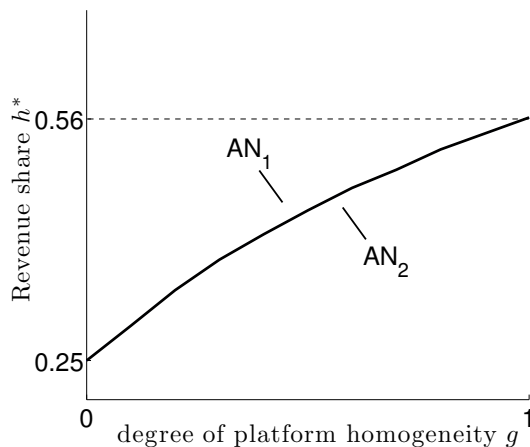


Figure 3: When symmetric, the ad networks select symmetric equilibrium prices.

THEOREM 2. *The game between the ad networks has a Nash equilibrium in pure strategies.*

PROOF SKETCH. The payoff functions of the ad networks are continuous and quasi-concave in a convex compact set. Thus, there exists a Nash equilibrium in pure strategies. \square

4. EXPERIMENTS

In this section, we gain some insights into the Nash equilibria of the game, through numerical experiments. We have shown in Section 3 that both ad networks will select an aggressiveness level of $x = 1$ and will compete in prices. Depending on the quality α_i of the classification algorithms of each network i , the estimated fraud intensity r , and the homogeneity g of the publishers, the two players adjust the revenue shares h they give out.

The first experiment studies the impact of the networks' efficiency in classifying valid clicks on the prices they give to the publishers. When the ad networks are of the same quality, the ad networks' NE prices are symmetric, as shown in Fig. 3. On the contrary, when the ad networks are asymmetric, we observe that the inferior network (AN_2 in our case) selects to give more to the publishers, as seen in Fig. 4.

We also explore the role of the publishers' homogeneity g in determining the prices in equilibrium. As g increases, the networks become more homogeneous, and the gap between the players' equilibrium prices increases (Fig. 4). When $g > 0$, there is a chance for one ad network to get all the market of publishers. Thus, we observe a fiercer competition on the revenue shares. The invalid fraction of clicks is $r = 0.3$, and the qualities of the ad networks are $\alpha_1 = 0.2$ and $\alpha_2 = 0.7$.

5. CONCLUSIONS

We presented a model to capture the incentives of ad networks to fight click fraud. The analysis shows that the ad networks maximize their revenues as the limit of the aggressiveness x of classification algorithms approaches 1. Therefore, the ad networks resort to competing in prices to attract a larger fraction of the publishers. Our results show that the more asymmetric in quality the ad networks are, the more asymmetric their equilibrium prices will be. Another finding of our work is that as the publishers become more heterogeneous, the competition in prices softens.

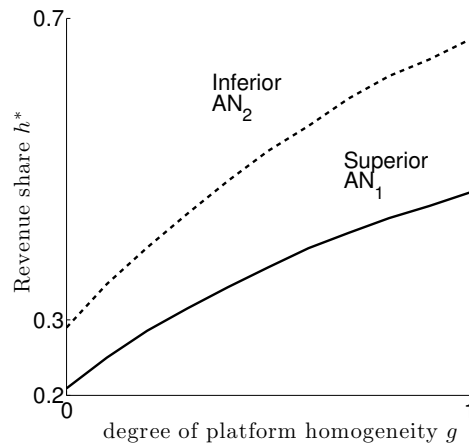


Figure 4: As publishers become more homogeneous, the inferior network (AN_2) offers a larger and larger share to publishers than the superior network.

6. REFERENCES

- [1] B. Bental and M. Spiegel. Network competition, product quality, and market coverage in the presence of network externalities. *Journal of Industrial Economics*, 43(2):197–208, 1995.
- [2] V. Dave, S. Guha, and Y. Zhang. Measuring and fingerprinting click-spam in ad networks. *SIGCOMM*, 42(4):175–186, Aug. 2012.
- [3] H. Hotelling. Stability and Competition. *Economic Journal*, 39(1):41–57, 1929.
- [4] J. Kim. The intensity of competition in the hotelling model: A new generalization and applications. MPRA Paper 6876, University Library of Munich, Germany, Apr. 2007.
- [5] B. Mungamuru, S. Weis, and H. Garcia-Molina. Should ad networks bother fighting click fraud? (yes, they should.). Technical Report 2008-24, Stanford InfoLab, July 2008.
- [6] J. Musacchio and D. Kim. Network platform competition in a two-sided market: Implications to the net neutrality issue. In *TPRC: Conference on Communication, Information, and Internet Policy*, Sept. 2009.
- [7] P. Njoroge, A. Ozdaglar, N. Stier-Moses, and G. Weintraub. Competition, market coverage, and quality choice in interconnected platforms. In *Workshop on The Economics of Networks, Systems, and Computation (NetEcon)*, Sept. 2009.
- [8] J. M. Perloff and S. C. Salop. Equilibrium with product differentiation. *Review of Economic Studies*, 52(1):107–20, 1985.
- [9] S. Sen, Y. Jin, R. Guérin, and K. Hosanagar. Modeling the dynamics of network technology adoption and the role of converters. *IEEE/ACM Trans. Netw.*, 18(6):1793–1805, 2010.