Economics of cookie matching (Final Project in Computational Games Theory Course)

Alex Fonar
I.D. 323301515
Slava Novgorodov
I.D.323624437

1 Introduction

In many cases, sites try to get more revenue from advertising. One way to do it is cookie matching. Cookies are small files placed on a user's computer that permit a website to record information about a previous visit. Website saves copy of this information too.

Cookies can be shared between sites. This helps to get more information about user, and so can be used to target advertising. This is most noticeable to the user when, for instance, they search for flights to Hawaii on Orbitz.com, and find Hawaii ads following them across many the web, for example showing up on the New York Times website.

It may seems, that cookie matching can only increase revenue from advertising. But it can cause revenue loss for some sites. This situation called information leakage. Suppose two sites: site "The Marker" newspaper, and site of "Israel Hayom" newspaper. "The Marker" site visited only by wealthy users, and "Israel Hayom" site visited by wealthy and poor users alike. And so, there are two advertisers: "Ferrari" seeking only wealthy users, and "Yossi Falafel" seeking only poor. Also there is 50% of wealthy people in society, and 50% of poor (all this information is only for the example purposes, and we suppose, is not quite true).

Suppose situation without cookie matching. "The Marker" site demands 10 cents for each user, seeing advertisement. "Israel Hayom" demands only 5 cents, because with 50% probability, advertisement will not go to right users.

Now, suppose there is cookie matching. Obviously it isn't help to "The Marker" site to distinguish between users, because site already knows, that everybody visiting it, is wealthy. But "Israel Hayom" can now distinguish between poor and wealthy users, and so take full price for advertising. Now, "Israel Hayom" propose price of 7 cent for user, seeing advertisement. "Ferrari" will prefer to advertise on "Israel Hayom" site, rather than "The Marker" site, and that will lead to revenue loss of "The Marker".

Let's assume a model where every user has the same value for all advertisers. In our example, it means that "Ferrari" and "Yossi Falafel", both value only wealthy users, and ignore poor ones. In the work, that we have been based "To match or not to match:

Economics of cookie matching in online advertising" by Ghosh,Madhian,McAfee and Vassilvitsky next statement was proved:

"When advertisers have identical rankings of users, publishers agree whether or not to share cookies. That is, either they all want to, or none want to, share"

And that is immediate conclusion from following theorem:

Theorem 1.1. *In both with and without cookie-matching models (with homogeneous advertisers), the expected revenue per impression of a website w is proportional to*

$$\beta_w := \frac{\sum_t p_t p_{t,w} \gamma_t}{\sum_t p_t p_{t,w}}$$

Therefore, either for all websites w_i, the revenue per impression of w_i in the model with cookie-matching is greater than its revenue per impression in the model without cookiematching, or the reverse inequality holds for all w_i.

Where p_t is probability that some user is of type t, $p_{t,w}$ is probability that user of type t will visit site w, and γ_t is probability of purchasing the product by user of type t.

Actually this model isn't frequent in real life, because almost in all cases different advertisers seeking different users.

2 Problem definition and our attempts for solution

2.1 Additional revenue partition

One of ways to solve information leakage problem is dividing of additional revenue got as result of cookie matching between sites.

Once again we return to example from introduction. Suppose "Ferrari" seeking only wealthy users, and "Yossi Falafel" seeking only poor ones. Also suppose that all Internet contain only 200 users: 100 wealthy and 100 poor. Wealthy users visit "The Marker" site and all users visit "Israel Hayom" site. And we assume that "Ferrari" tries to get as much attention as possible, and than tries to advertise on both sites. But each site has place for only one advertisement.

In model without cookie matching, "The Marker" site demands 10 cents for each user, seeing advertisement. So, expected income of "The Marker" is 10*100 = 1000 cents. "Israel Hayom" gets only 5 cents per user, no matter which advertisement is used, and has expected income of 5*200 = 1000 cents.

In model with cookie matching, "The Marker" still has expected income of 1000 cents. From the other hand now "Israel Hayom" knows which users visiting it are wealthy, and which are poor. So it can show "Ferrari" advertisement to wealthy users, and "Yossi Falafel" advertisement to poor ones. In that case site gets 10 cents per each user, and has expected income of 10*200 = 2000 cents.

Now there is 1000 cents of additional income. All of that income was earned by "Israel Hayom" site, but it has not been possible to get without "The Marker". Actually, "The Marker" site has no interest in cookie matching at all, and can stop giving cookie information to "Israel Hayom". So, this additional income of 1000 cents must be divided in some way between 2 sites.

In general case, all additional revenue that was earned as result of cookie matching, must be divided in some way between sites. The problem is how to divide it. In some cases we can use network exchange theory to solve this problem.

Information about Nash bargaining solution and network exchange theory was obtained from "Balanced Outcomes in Social Exchange Networks" work by Kleinberg and Tardos.

2.2 Nash bargaining solution

I will start from most basic bargaining example. Suppose that two parties are negotiating over how to divide one unit of money, and that each has an alternate option - a fallback amount that it can collect in case negotiations break down without a division. How this amount of money will be divided? There is Nash bargaining solution for this problem: Suppose agents A and B have alternate options alpha and beta respectively. Then the Nash bargaining solution posits that they split the surplus $s = 1 - \alpha - \beta$ evenly between them: if $s < 0$ then there is no solution that will make them both happy, while if $s \geq 0$ then they will agree on $\alpha + s/2$ for A and $\beta + s/2$ for B.

2.3 Network exchange theory

Here is a statement of a problem. In a controlled laboratory setting, n human subjects are asked to each play the role of one of the n nodes of a graph G. The value in each social relation is captured by placing a fixed sum of money (say, \$1) on each edge. Nodes now engage in free-form negotiation (say, via instant messaging from different computer terminals) over how to divide the units of money on each edge. By the end of the time limit, each node v is supposed to reach an agreement with at most one neighbor w on how to divide the money on their incident edge; if v is able to reach such an agreement, then she gets her agreed upon share of the money on this edge, and if v is not able to reach an agreement with any neighbor, then she gets 0. Some experiments were done to check result of bargaining of that type. And there are standard methods of prediction of outcome in many cases (such as balanced outcomes and Klenberg & Tardos work). For example in case of most simple graph containing only 2 nodes and edge between them (Fig. 1), the prediction is that 2 participators will reach a split close to $\frac{1}{2} - \frac{1}{2}$.

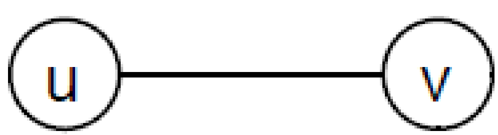

Figure 1: 2 - node network

In "The Marker"/"Israel Hayom" example additional 1000 cents are earned only in case of 2 sites cooperation, so we can state that eventually 2 sites will divide income close to 500 - 500.

In other example of 3-node network (Fig. 2) middle node has very powerful position. Since v can only complete one successful transaction, one of nodes u or w in this network will necessarily be left out and hence get 0. Node v uses this power over u and w to obtain close to the full amount on the edge where she does reach an agreement, and so even the node among u and w that completes a transaction gets close to 0.

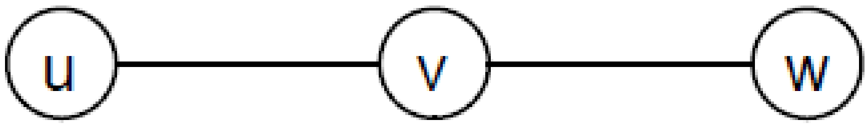

Figure 2: 3 - node network

In 5-node network from Fig.3, v still has very powerful position as in previous example, and will reach an agreement with u or w. Meanwhile, node x quickly realizes that it's useless to bargain with v, and so he spends most of his energy bargaining with y on essentially equal terms, reaching a split close to $\frac{1}{2} - \frac{1}{2}$.

Also same problem can be stated and in some cases resolved in weighted graphs, where value being divided on each edge can be arbitrary positive number.

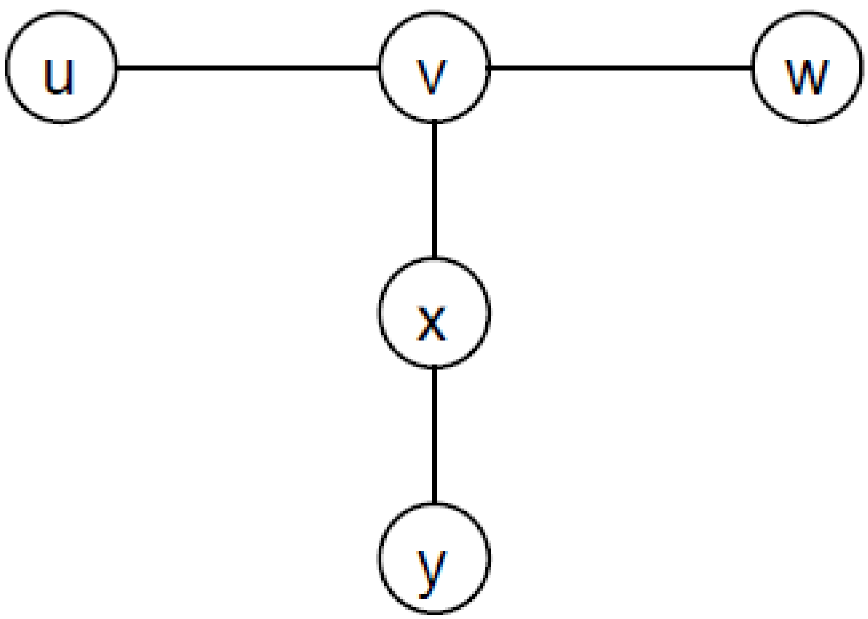

Figure 3: 5 - node network

2.4 Using of network exchange theory for additional income partition problem

In some cases problem of dividing income between sites can be simplified into network exchange problem. For example assume there is 3 sites: "The Marker", "Globes" and "Israel Hayom". "The Marker" and "Globes" have exactly same auditory. So in this case, we can use network from Fig. 2, when "Israel Hayom" occupies powerful middle node, and can choose site for creating cookie matching agreement. In this case "Israel Hayom" will obtain almost all additional income.

But it seems that such approach to additional revenue partition problem can be used only in very small amount of cases, where are no cookie matching agreements between 3 or more sites. In fact, network exchange theory cannot be used even if there are only 3 sites, and there was some additional income from crossing cookie information from all of them. So we tried to resolve problem for this case.

2.5 Additional revenue problem in case of 3 sites

2.5.1 Problem statement

There are 3 sites. And there are 5 possibilities:

1. No agreement is reached. All sites get 0.
2. Sites 1 and 3 reached cookie matching agreement, and will divide a\$ between them.
3. Sites 2 and 3 reached cookie matching agreement, and will divide b\$ between them.
4. Sites 1 and 2 reached cookie matching agreement, and will divide c\$ between them.
5. All 3 sites reached cookie matching agreement, and will divide D\$ $(D > a, b, c)$

Without possibility 5(marked as P5 below), this is simple weighted network exchange problem (Fig. 4)

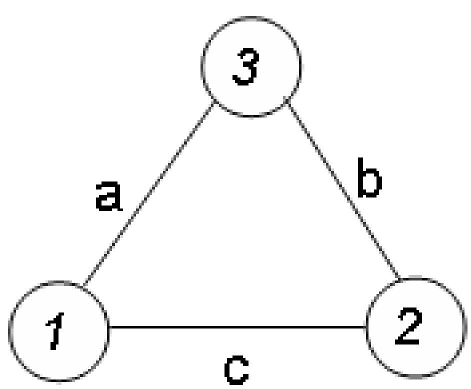

Figure 4: Triangular network

There are two cases:

2.5.2 Weight of some edge is not less than sum of 2 other edges weights. Say $c \geq a + b$.

We start from assumption that P5 (3 - sites agreement possibility) doesn't exist. In this case there is simple (and if $c \neq a, b$, unique) balanced outcome solution: Sites 1 and 2 will reach an agreement. Site 1 has alternate option of a, site 2 has alternate of b. So surplus is $s = c - a - b$, and will be divided equally between 2 sites. So site 1 will get $a + \frac{s}{2} = \frac{a+c-b}{2}$, and site 2 will get $b + \frac{s}{2} = \frac{a+c-b}{2}$. We will mark this allocation as $(\frac{a+c-b}{2}, \frac{a+c-b}{2}, 0)$

Now we return P5. We check what will be results of 3-sites agreement. Now this is Nash bargaining solution case. 2-site agreement between 1 and 2 is alternate options for these two sites. Site 3 has alternate option of 0. Surplus is $t = D - \frac{a+c-b}{2} - \frac{a+c-b}{2} = D - c$, and it will be equally divided between sites. So we get allocation $(\frac{a+c-b}{2} + \frac{t}{3}, \frac{a+c-b}{2} + \frac{t}{3}, \frac{t}{3}) = (\frac{3a+c-3b}{6} + \frac{D}{3}, \frac{3b+c-3a}{6} + \frac{D}{3}, \frac{D-c}{3})$ For example if $a = b = 20, c = 50$ and $D = 60$, sites 1 and 2 will get 28 1/3$ each, and site 3 will receive only 3 1/3$.

2.5.3 Every 2 edges weights greater than weight of the third edge.

If we drop P5, network exchange problem has no balanced outcome, and it seems that real outcome cannot be predicted without extra assumptions. We started from most simple case.

2.5.3.1 $a = b = c$

For simplicity of description we will assume that $a = 80$(on Fig. 5 you can see network in this case).

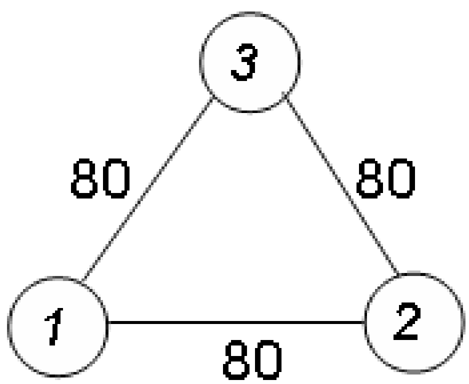

Figure 5

Once again, we start from assumption that P5 doesn't exist. From symmetry of the network, is obvious that there is no unique balanced outcome. Moreover, we even cannot say that in the end 2 players will reach (40,40) agreement. There is an example of one turn of events (example with slight changes is taken from Y. Narahari lectures on cooperative game theory):

- If sites 1 and 2 negotiate effectively in the coalition, they can agree to equally divide 80$ between them. We will mark this allocation by (40, 40, 0).

- If (40, 40, 0) is the expected outcome, then site 3 would be eager to persuade site 1 or site 2 to form an effective coalition with him. For example, site 3 would be willing to negotiate an agreement with site 2 to both propose (0, 60, 20).
- If (0, 60, 20) were to be the expected outcome in the absence of further negotiations, then site 1 would be willing to negotiate an agreement with site 3 to propose an allocation that is better for both of them, say, (30, 0, 50).

It turns out that in any equilibrium of this game, there is always at least one pair of sites who would both do strictly better by jointly agreeing to change their strategies together.

The above sequence of negotiations seems to have no end. But we still believe that 2 sites will reach (40,40) agreement. It seems that at first step 2 sites will negotiate on (40,40) agreement. Say, it sites 1 and 2. We saw that if one of those sites is greedy and try to get more income from reaching an agreement with site 3, we enter infinite loop of agreement proposals. If we add a condition that sites have limited time to reach an agreement, we still see that outcome cannot be predicted. So, if site 1 will try to reach agreement with site 3, with probability of 1/3, site 1 won't get nothing, and with probability of 2/3 will get some of money with expected value of 40$ (total expected value is quite predictable: 80/3 $). So site 1 is better to stay with initial (40,40) agreement with site 2. Same is right for site 2 as well. So it will end with (40,40,0) agreement.

But "proof" above has one problem: negotiations can start from point different than (40,40). Suppose one site tries to increase chances to take part in agreement, and start negotiate from point of demanding only 30$ from a deal. Suppose it site 1. Site 2 will be on the point of reaching agreement with site 1 on terms of (30,50,0), when site 3 will interrupt it, and propose to site 1 agreement of (32,0,48). If site 1 accepts, site 2 will propose counter agreement to site 1(and not to site 3, which will demand more money). So if site 1 is on point of receiving of agreements with site X(2 or 3), in which site 1 gets less than 40$, it will receive proposal with better offer from remaining site. But this is only until site 1 getting less than of 40$. You can see that as long as site 1 demanding less than it partner, site 1 is on less vulnerable position, because site 1 is the address of next counterproposal of the remaining site. Now we define a term of *vulnerability.*

Definition 2.1. In triangular network of sites i, j, k, where agreement between sites i and j is about to take place, site i is vulnerable if next counterproposal of site k will be to site j.

If edge weights on network are equal, only way to avoid *vulnerability*, is to demand less than your partner. So if we state a new assumption that every site is trying to avoid *vulnerability*, the only possible balanced situation is (40,40) agreement between some 2 sites.

We will end this line of not very scientific arguments here, and simply suppose, that original problem (without P5) will end with $(a/2, a/2)$ agreement between some 2 sites(say 1 and 2). Now we return P5. The next seems quite simple. We have standard case of

Nash bargaining. Sites 1 and 2 have alternate options of $\frac{a}{2}$. Site 3 has alternate option of 0. So surplus is $s = D - a$, and it will be divided equally between sites, also sites 1 and 2 will receive their alternate options. Sites 1 and 2 get $\frac{s}{3} + \frac{a}{2}$ each, site 3 only $\frac{s}{3}$. If $a = 80$ and $D = 90$, we get allocation of (43 1/3, 43 1/3, 3 1/3). It seems reasonable because equal money partition in 2-side agreement gives more revenue than equal partition in 3-side agreement. But suppose $a = 80$ and $D = 180$. From our formula we get allocation of (73 1/3, 73 1/3, 33 1/3). But it seems, that as result of network symmetry and 3-side agreement obvious advantage on 2-side agreement, right allocation is (60,60,60). We think, that both allocations can occur. (73 1/3, 73 1/3, 33 1/3) can occur when 2 players form coalition, and "rob" third player. (60,60,60) can occur, if 3 players reach an agreement from the start.

2.5.3.2 General case (Fig. 6)

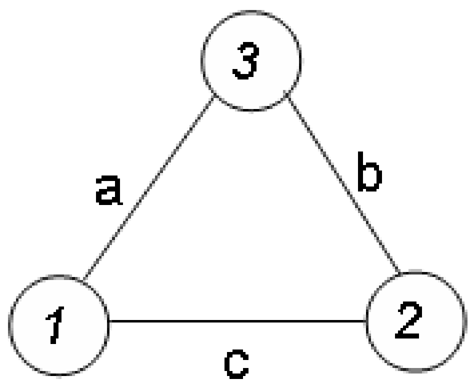

Figure 6

Once again we start from assumption that P5 doesn't exist. And as in previous section we assume that every player tries to avoid *vulnerability*.

I will start from simple example (Fig. 7)

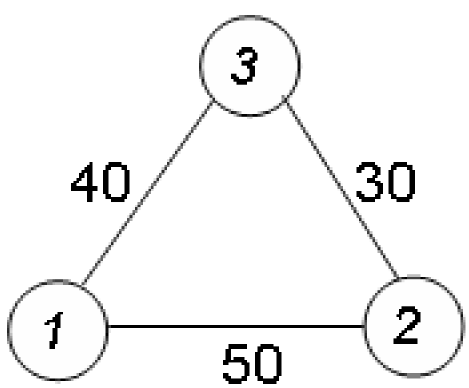

Figure 7

Suppose sites 1 and 2 are about to "sign" an agreement. For example, they agree on (25,25,0) allocation. Now if site 3 suggest counterproposal, it must offer at least 25$, leaving to site 3 no more than 15$ in case of proposal to site 1, and no more than 5$ in case of proposal to site 2. So site 3 will prefer to make an agreement with site 1 We are seeking some sort of equilibrium, so that site 3 doesn't have a preferable site for counter proposal. In this case it is initial (30,20,0) agreement between sites 1 and 2. No matter if site 3 counter propose to site 1 or site 2, site 3 can reckon on no more than 10$.

Now we solve this equilibrium problem in general case (Fig. 6). There are 3 cases:

- Sites 1 and 2 are about to "sign" an agreement. We try to find terms of such an agreement that form equilibrium. Suppose they agreed on $(x, y, 0)$, when $x + y = c$. Site 3 can reckon on no more than $a - x$ in case it counter propose to site 1, and on no more than $b - y$ in case of proposal to site 2. In equilibrium:
  $a - x = b - y$
  We know that $y = a - x$, and so we get $x = \frac{a+c-b}{2}$ and $y = \frac{b+c-a}{2}$.
- Sites 1 and 3 are about to "sign" an agreement $(x, 0, z)$. In this case we get $x = \frac{a+c-b}{2}$, $z = \frac{a+b-c}{2}$.
- Sites 2 and 3 are about to "sign" an agreement $(0, y, z)$. In this case we get $y = \frac{b+c-a}{2}$, $z = \frac{a+b-c}{2}$.

You can see that we got same numbers assigned to particular site. Only in each case there are only 2 sites, that get assigned amount of money, and third one get nothing. The assigned amount of money is always (sum of weights of 2 adjacent edges minus weight of opposite edge)/2. In case that weight of some edge is not less than sum of 2 other edges weights, one of that assigned numbers is negative or 0, and we get balanced outcome solution.

And for the end, we return P5. As in previous section, we assume that initially 2 sites form a coalition that in 3-side agreement takes majority of income, leaving only third of surplus to third site. If sites 1 and 2 reached initial agreement of $(x, y, 0)$ final agreement will be $(x + \frac{D-c}{3}, y + \frac{D-c}{3}, \frac{D-c}{3})$. Similar situation occur in 2 other cases. P5 changes the equilibrium a bit, and in new situation it will be:

- If sites 1 and 2 form a coalition, and site 3 counter propose to site 1, revenue that sites 1 and 3 can get equal to: $\frac{2(D-a)}{3} + a$. The minimal revenue of site 1 is $x + \frac{D-c}{3}$. So site 3 can reckon on no more than
  $\frac{2(D-a)}{3} + a - (x + \frac{D-c}{3})$.
  Symetrically in case of proposal to site 2, site 3 can reckon on no more than
  $\frac{2(D-b)}{3} + b - (y + \frac{D-c}{3})$.

  If we are seeking equilibrium this 2 expressions must are equal. So, we get
  $x = \frac{a+3c-b}{6}$, $y = \frac{b+3c-a}{6}$, and allocation will be:
  $(\frac{a+c-b}{6} + \frac{D}{3}, \frac{b+c-a}{6} + \frac{D}{3}, \frac{D-c}{3})$

- If sites 1 and 3 form a coalition, allocation will be
$(\frac{a+c-b}{6} + \frac{D}{3}, \frac{D-a}{3}, \frac{a+b-c}{6} + \frac{D}{3})$

- If sites 2 and 3 form a coalition, allocation will be
$(\frac{D-b}{3}, \frac{b+c-a}{6} + \frac{D}{3}, \frac{a+b-c}{6} + \frac{D}{3})$

Once again we get symmetric allocation in 3 cases, with amounts of money assigned to each site. In each case there are only 2 sites, that get that assigned sum, and third one get only relatively small value . The problem is that result above does't agree with result from subsection 2.5.2, which was $(\frac{3a+c-3b}{6} + \frac{D}{3}, \frac{3b+c-3a}{6} + \frac{D}{3}, \frac{D-c}{3})$, for sites 1 and 2 agreement. Maybe, simply both results can be possible in different situations.

3 References

- Ghosh A., Mahdian M., McAfee R. P., Vassilvitskii S. 2012. To match or not to match: Economics of cookie matching in online advertising

- Kleinberg J., Tardos E. 2008. Balanced Outcomes in Social Exchange Networks.

- Narahari Y. 2012. Lecture notes in Cooperative Game Theory, Coalitional Games: Introduction