

Consolidated Review of

Follow the Money: Understanding Economics of Online Aggregation and Advertising

1. Strengths

The paper makes several relevant findings that advance the community's knowledge of a topical subject. The findings pertain to:

1. Classifying various internet users and determining the properties of users that generate the most revenue,
2. Evaluating publisher profits and determining the properties that makes one publisher more profitable.
3. Determining the impact of users' taking action to prevent aggregators from collecting information -- which revenue declines by a large fraction. All finds are extremely useful, the third shows that aggregators have a vested interest to collect data and would not willing participate in schemes that reduce the amount of information collected. The first two can help community as we attempt to model various aspect of the online advertising ecosystem.

It is a nice analysis and, so far as I know, the first at this level of detail. Well documented as to how they picked their various parameters that they then drive with their data sets

2. Weaknesses

The bulk of the results are not particularly surprising (but that's a subjective opinion).

Most of the paper focuses on results from one data set, it isn't clear how generalizable several of the findings. Also the data set focused on, namely mobile users, may introduce a significant bias when compared to other data sets, as usage patterns are significantly different.

TQM is, as best I can tell, a fudge factor (indeed, 3.4 says the paper sought "trends... that are not impacted by scalar values such as TQM") yet it seems to me that tables 2 and 3, given they reflect estimated revenue (which is computed using TQM) are not fudge free. The language is occasionally difficult to understand.

3. Comments

I don't have any technical criticisms. There are only a couple of things that -- as a non-expert -- I would have liked to see explained in a bit more detail:

- ❖ It is not 100% clear whether the model represented by Equation 1 is novel or not. At first I thought that writing down this equation was part of the paper's contribution. But then it turned out that the terms that make up the equation have been defined before and have known value ranges. I.e., if RON, TQM, and user intent are previously-defined terms and their values are taken from existing text, what is the contribution of the paper in terms of modeling?
- ❖ It seems that the results of the paper depend significantly on the intent values for categories. Hence, I think the authors should spend more time explaining how these are determined and perhaps also show some example values for a few basic categories. E.g., suppose we have assigned bid value 2 to "sports" and 3 to "pets". What about the "sports + pets" category? Does it get a 5? Probably not, because -- I am

guessing -- there may be an overlap between the kinds of information we get about a user based on his/her interest in each of the two categories.

- ❖ In Section 4.2, it says, "we find that some categories are more lucrative than others." Is this really something that the authors "find", or something that they derive directly from the intent values for categories given by Adwords?
- ❖ Why exactly does Google generate less revenue than Facebook as a publisher? Because there are fewer aggregators present on YouTube?

Generally a good paper and a fun read. Comments as I found them.

- ❖ Intro "Facebook is increasing their presence" -- this may be a regionalism, but I prefer "its presence"
- ❖ Intro "(b) few aggregators" -> "(b) a few aggregators".
- ❖ In section 2.3, equation 1. The equation has three values multiplied together. RON is a price that you can find (or at least estimate). User Intent is also well-defined. But TQM seems poorly grounded. The willingness of advertisers to pay for intent (2-10X) is documented in 3.3, but that's part of User Intent. So what factor does TQM measure? It seems to be a tool to simply delete junk publishers from the data set, but that's not how it is described here.
- ❖ Section 3.2, "common aggregators that emphasizes" -> "common aggregators. That emphasizes"
- ❖ Section 3.4, fine to say TQM covers other factors, and that full coverage is beyond scope, but you need to at least list those factors -- giving just one example (ad placement) is too cavalier.
- ❖ Section 4.1, the fact that many users only hit a small number of sites (and thus, $EI(u)$ can be excellent with partial coverage) also suggests that aggregators may be missing some major targets. For instance, imagine that smaller spenders hit more sites (perhaps because they wish to look more places before making a buy decision) or that there are specialty spenders who hit specialty sites. I'm not asking you to solve this question -- but a quick observation that there's a research issue hiding here, namely do aggregators, due to their limited coverage, completely miss valuable niche markets?

The paper focuses on a mobile data set, given this it would be interesting to further subdivide to android and iPhone and try to understand how this impacts things; I suspect that there could be a huge impact. I believe android mobile phones have many more services that tie them directly into the Google ecosystem. It would also be interesting to state if the observations are generalizable to other data sets. This is done in some subsection but not all.

Outside Reviewer: This is a very interesting article. In response to your question (insofar as I understand it), the theoretical basis for the analysis looks very sound. If you are asking how it relates to conventional media planning, it tracks very well. My only

criticism is that it seeks to reinvent the planning process out of whole cloth rather than building on the conventional media literature, then showing how this applies to the principles in the data-rich Internet environment. Of course, the article is not about media planning, but rather, the financial consequences of data availability. But the analysis is based on an analog to standard media planning concepts. I have attached a paper* that summarizes the conventional model to which I allude. $CPM(u,p,a)$ in the ACM submission corresponds roughly with $CPMETM$ in the attached paper. The ACM paper would have to rationalize the various players in the Internet space (publishers, aggregators, user-agents) against the conventional structure, but this would be simply to do and would provide a good bridge in the media literature. I hope this addresses what you were asking for. Thanks for sharing the paper. I found it very interesting, as should the folks at ACM. [*PC co-chair's note -- the paper is by Hugh Cannon, "Addressing New Media with Conventional Media Planning", from the Journal of Interactive Advertising, Vol 1, No. 2.]

4. Summary from PC Discussion

The PC discussion was short. The issue up until the PC meeting was concern about reviewer expertise. We got an outside reviewer (from the on-line advertising community) who validated that the paper's contribution was worthy of acceptance and so the paper was accepted.

5. Authors' Response

First of all, we are thankful to the reviewers for great feedback and to our shepherd, Craig Partridge.

There seem to be two (relatively) large concerns: the first deals with the parameter TQM, and the second corresponds to related work in the advertising literature.

Regarding the first concern, we used TQM (traffic quality multiplier) as a parameter to represent a factor like quality of the publisher after conversing with advertising professionals. Top publishers who attract more eyeballs or have a 'brand prestige' factor increase the revenue for advertisers. Different ad-networks have different methods to assign values to this factor hence we issued a disclaimer about not giving serious consideration to absolute values.

We did learn that publishers that are deemed 'spam' are highly undesirable to advertise on and hence, we assigned a low value to them (0.1). For top 500 publishers and non-spam publishers, we assigned 2 and 1, hence the differential factor is a constant, as opposed to an order of magnitude difference between spam (0.1) and other publishers (1 or 2). We have clarified this point at different places in the paper.

Regarding the second concern, we delved into the Journal of Interactive advertising and did not find much directly related work. What little we found, we mentioned in the related work section.