Measurement and Analysis of OSN Ad Auctions

Yabing Liu
Northeastern University
Boston, MA
ybliu@ccs.neu.edu

Chloe Kliman-Silver
Brown University
Providence, RI
chloe.klimansilver@gmail.com

Robert Bell
AT&T Labs–Research
Florham Park, NJ
rbell@research.att.com

Balachander Krishnamurthy
AT&T Labs–Research
Florham Park, NJ
bala@research.att.com

Alan Mislove
Northeastern University
Boston, MA
amislove@ccs.neu.edu

ABSTRACT

Advertising is ubiquitous on the Web; numerous ad networks serve billions of ads daily via keyword or search term auctions. Recently, online social networks (OSNs) such as Facebook have created site-specific ad services that differ from traditional ad networks by letting advertisers bid on *users* rather than *keywords*. With Facebook's annual ad revenue exceeding $4 billion, OSN-based ad services are emerging to be a significant fraction of the online ad market. In contrast to other online ad markets (e.g., Google's ad market), there has been little academic study of OSN ad services, and OSNs have released very little data about their advertising markets; as a result, researchers currently lack the tools to measure and understand these markets.

In this paper, our goal is to bring visibility to OSN ad markets, focusing on Facebook. We demonstrate that the (undocumented) feature that suggests bids to advertisers is most likely calculated via sampling recent winning bids. We then show how this feature can be used to explore the relative value of different user demographics and the overall stability of the advertising market. Through the exploration of suggested bid data for different demographics, we find dramatic differences in prices paid across different user interests and locations. Finally, we show that the ad market shows long-term variability, suggesting that OSN ad services have yet to mature.

Categories and Subject Descriptors

J.4 [**Social and Behavioral Sciences**]: Economics; H.3.5 [**Information Storage and Retrieval**]: Online Information Services—*Web-based services*; J.4 [**Social and Behavioral Sciences**]: Sociology

General Terms

Measurement, Experimentation, Economics

Keywords

Online advertising; Online social networks; User demographics; Suggested bid

1. INTRODUCTION

Advertising is now the economic underpinning of much of the Web; large advertising networks (e.g., Google's Ad Network [22]) serve advertisements for millions of Web sites. Many of these advertising services are implemented as auctions, with individual advertisers bidding on specific keywords, pages, or search terms. These auctions are extremely popular with advertisers (Google alone earned over $50 billion in advertising revenue in 2013 [21]) and are well-studied in the research literature [15, 33, 37, 39].

Recently, a new type of advertising network has emerged [24]: closed-site advertising services run by online social networks (OSNs) such as Facebook. Unlike prior systems, where the advertising network was forced to infer user information from cookies, browsing history, and search terms, OSN-based advertising services are provided demographic information directly by the users themselves. As a result, advertisers are able to target *users* directly (via profile attributes), rather than targeting keywords or search terms. Although OSN-based advertising services are nascent, they already carry a significant number of ads: Facebook alone had over $7.8 billion in advertising revenue in 2013 [10].

Unfortunately, there has been little academic study of these ad networks, and OSNs have released very little data about their advertising markets; the most in-depth numbers are from U.S. Securities and Exchange Commission (SEC) filings by the OSNs, which are typically at per-continent-per-fiscal-quarter granularity. Thus, researchers have little visibility into the dynamics of these markets, and it remains unclear which user demographics are the most valuable to advertisers (and therefore to the OSNs) and how stable these values are over time.

In this paper, our goal is to develop techniques that will allow researchers to measure and understand OSN ad markets. We focus on Facebook (currently the largest OSN ad market) and make three contributions: First, we explore how the *suggested bid*—a common feature of ad services that suggests prices to bid for a given target demographic—can provide insights on the revenue attainable from different users. On Facebook, the suggested bid is an undocumented feature, and the internal algorithm that Facebook uses is not public.

Basic Fields	Parameters/Examples
Location	Country, State, City, Postal code
Gender	Male, Female, All
Age	Range (from 13–65)
Precise Interest[1]	Travel, Science, Music, ...
Broad Category[2]	Cooking, Gardening, iPhone 5, ...
Interested In	Male, Female, All
Relationship Status	All, Single, In a relationship, Married, Engaged, Not specified
Language	English, Spanish, French, ...
Education	Anyone, In high school, In College, College Grad
Workplaces	Google, Facebook, AT&T, ...

Table 1: Facebook's targeting parameters made available to advertisers.

However, we demonstrate that this feature is likely based on a sample of the recent winning bids on users in the target demographic, and we provide strong supporting evidence for this hypothesis by conducting an experiment where we actively participate in the ad market.

Second, we demonstrate how researchers can use the suggested bid data. The raw data returned from the queries are noisy due to the sampling methodology, but we demonstrate that repeated sampling of the ad market can provide consistent results with distinctive trends. We verify that our derived relative revenue per user correlates well with ground-truth figures from Facebook's SEC filings [14]. While our methodology focuses on Facebook, it likely can be applied to other OSNs that provide suggested bids for placing advertisements. We make all of our code and data available to the research community.

Third, we use the suggested bid mechanism to explore two questions about how different users contribute to Facebook's revenue: How do the advertising prices compare across different demographics, and how stable are the prices for different target demographics over time? We explore different attributes, including location, age, and user interests, and provide a summary of the distribution of prices paid to advertise to different user demographics. We find significant differences in ad prices across different locations and user interests, and fewer differences by age. We also find prices to be variable over the long-term, but with distinct trends; this is consistent with OSN-based advertising markets being in a nascent phase.

The rest of this paper is organized as follows: Section 2 provides background information on Facebook's advertising model and an overview of related work that studies online advertising. Section 3 describes our data collection methodology, examines the properties of the suggested bid data, and explores how it can be interpreted and used. Section 4 presents an analysis of the current Facebook ad market, exploring the prices of different demographics and the stability over time. Finally, Section 5 provides a concluding discussion.

---

[1]Precise interests are interests explicitly stated by the user in their profile; broad interests are inferred by Facebook based on user activity.

[2]Broad Categories are pre-defined targeting categories provided by Facebook that group users according to their Likes, interests, applications, and other profile content they have provided. Recently, Facebook changed the interface by

2. BACKGROUND

We now provide background on Internet advertising and detail related work on measuring online advertising auctions.

2.1 Online advertising

Most online advertising today is placed via *auctions*, where advertisers bid on keywords, search terms, or (in the case of OSNs) user demographics. The advertising network selects the winning bidders and presents their ads to the users. In different online advertising platforms, the underlying auction mechanism varies: many traditional ad networks, such as Google and Yahoo, use Generalized Second Price (GSP) auctions [9], whereas Facebook uses Vickrey–Clarke–Groves (VCG) auctions [36]. These two mechanisms primarily differ in how they calculate the price that a winning bidder should pay.

Advertisers typically bid using either CPM (Cost Per Mille, the cost of 1,000 ad impressions) or CPC (Cost Per Click). In order to support both bidding mechanisms at once, ad networks will typically record each advertiser's click-through rate (CTR, the fraction of impressions that result in a click), providing a way to compute an estimated CPM bid given a CPC bid (i.e., the estimated CPM bid is simply the CPC bid multiplied by the advertiser's CTR). In this paper, we focus only on CPM ad prices, as reasoning about CPC requires knowing an advertiser's CTR (which is not always available).

2.2 Facebook advertising

Advertisers can place ads on Facebook by creating *campaigns*; each campaign consists of a specific ad, a target demographic, a CPC or CPM bid, and a budget. After an advertiser creates a campaign, it must first be approved by Facebook (this process typically takes less than one day). Once the ad campaign is active, the advertiser participates in the ad auctions that occur whenever a user in the target demographic is shown an ad. An advertiser can view their campaign's status on Facebook's Web site to see details on the number of impressions, clicks, unique users, and overall cost, and can pause or cancel their campaign at any time.

Unlike traditional ad networks, OSNs such as Facebook have significant data about each user, including their personal information (demographics, interests, educational history, relationship status, etc), identities of friends, and their activity on the OSN. An overview of the currently available

---

merging existing categories into a more general "Interests and Behaviors".

Figure 1: Screenshot of Facebook's ad creation webpage, showing the suggested bid (bottom right) for the selected targeting parameters. We programmatically collect the suggested bid information.

targeting parameters[3] are shown in Table 1. Advertisers can target any combination of these parameters, and are only required to specify at least one country.[4]

2.3 Related work

There is surprisingly little research work on OSN ad auctions. Researchers have studied existing Web-search-based advertising networks (e.g., Google's ad market) [15, 32, 39], prediction markets run by Google, Ford and others [5], and have used "estimated prices" [17] from Google's Traffic Estimator Tool [23] (similar to our suggested bids) as a mechanism for understanding the network. Noti et al. [29] demonstrated that bidders with some explicit knowledge during an initial learning phase can bid with better valuations than those without such knowledge; this implies that suggested bids can provide useful guidance for bidders. A related study [11] has developed a new analytical model of GSP auctions in order to predict the number of clicks and the total price the advertiser can expect, using the advertiser's bid and the distribution of the number of opponents and their relative weighted bids. Our work provides another methodology for the advertiser to estimate the price they need to pay when targeting users with different demographics in OSNs, in which the information of opponents are not available. Furthermore, we provide a way for advertisers to estimate and compare the prices for targeting different demographics of users in an automatic way; previously, advertisers needed to manually select a combination of interests, age, gender, and other options to obtain Facebook's suggestions.

Researchers have also examined and improved auction mechanisms, including better CTR estimates [6, 16], usage of reserve prices in ad auction [30, 34], and optimization in

multiplicative bidding [3]. Using a sample from a week's worth of data across all keywords on Bing, other work [4] showed how to optimize linear combinations of the stakeholder utilities, showing that these can be tackled through a GSP auction with a per-click reserve price. There has also been much work proposing new models for conducting online auctions [19, 20, 28, 39].

Prior work [18] has shown that the contribution of users to advertising revenue is skewed (20% of users accounting for 80% of revenue), which provides supports for our observations that the OSN ad market prices vary widely across different user interests and locations. Another study [2] has utilized AdReveal, a browser based tool to provide measurements of 139K online display ads and analysis of 103K Web pages. They demonstrated that up to 65% of ad categories received by users are behaviorally targeted using users' online interests. Our study has shown similar result: targeting users with specified interests is more expensive, implying more competition among advertisers.

Much of this work is orthogonal to ours, as we present a mechanism for measuring the OSN ad network itself. Others have explored ways of identifying influential users in OSNs [1, 8]; our work is complementary to these. Additionally, there are nascent systems for leveraging unique features of OSNs for advertising (e.g., adby.me [25] allows users to create their own ads, shown to their friends). Finally, a few companies (e.g., AdParlor) have used Facebook's suggested bid data to provide clients with information on the ad market. Unfortunately, there are few published details of how their analysis is conducted.

3. SUGGESTED BIDS

We now describe our approach for measuring the Facebook Advertising Platform using suggested bids. We analyze the properties of the suggested bid data, with a goal of determining what the suggested bids represent and how they are calculated by Facebook. Finally, we describe how we interpret and use the data for our analysis in the following section.

---

[3]https://developers.facebook.com/docs/reference/ads-api/targeting-specs/
[4]Facebook provides different granularities of location information in different countries: the US includes ZIP code, most western European countries include city/town, and most developing countries only allow targeting at the country level.

Query:

```
https://graph.facebook.com/reachestimate?targeting_spec=
{"countries":["US"],"age_min":21,"age_max":30,genders=[1]}
&currency=USD&accountId=XXX&access_token=XXXX
```

Response:

```
{"data": {"users":62984500,"bid_estimations":
[{"location":3,"cpc_min":54,"cpc_median":82,"cpc_max":144,
"cpm_min":3,"cpm_median":14,"cpm_max":83}]}}
```

Figure 2: Example query targeting U.S. males between 21 and 30. Also shown is the suggested bid response (in JSON format).

3.1 Collecting suggested bids

Facebook provides an API[5] that allows advertisers to create and manage ads efficiently by making different API queries. Unfortunately, the Ads API is restricted to high-volume advertisers and we were unable to obtain access. Instead, we obtain suggested bid data from Facebook's Ad Creation Web page,[6] a screenshot of which is shown in Figure 1.

We programmatically send HTTP GET requests to the Facebook URL that serves the suggested bid data presented on the page. This URL accepts arguments representing the desired targeting parameters and returns a suggested bid to the user. A user must be logged in to make the suggested bid request, so we create a pool of three accounts to place the queries. These accounts have no prior advertising history or uploaded content.[7]

The suggested bid response from Facebook actually contains 7 different values: the estimated *audience size*[8] (i.e., the number of Facebook users in the target demographic), the CPC min/med/max, and the CPM min/med/max. An example of query and response data is shown in Figure 2.[9] It is important to note that all suggested bid data is public for all advertisers, and contains no personally identifiable information for either the other advertisers or other Facebook users.

3.2 Suggested bid observations

Given the suggested bid data made available by Facebook, we now provide five observations about the bid data. To help illustrate these observations, we collect a data set consisting of 1,000 suggested bids in quick succession (i.e., issuing 1,000 queries back-to-back within 30 seconds), separately targeting each of the 204 countries that Facebook supports. Thus, for each country, we collect a set of 1,000 suggested bids obtained in quick succession, where each bid uses the default targeting parameters with the exception of specifying the re-

---

[5]https://developers.facebook.com/docs/ads-api/

[6]https://www.facebook.com/advertising/

[7]The advertising history of accounts could be a factor in determining the suggested prices. To control for this, we tested with different accounts and observed no significant differences in characteristics of the returned suggested bid data.

[8]The audience size is returned with a granularity of 20 users, presumably to provide user privacy.

[9]The location parameter that is provided in the result is undocumented, but likely to the different locations on Facebook in which the ad will be shown.

	Coefficient of Variation		
Country	CPM min	CPM med	CPM max
US	0.176	0.143	0.260
NZ	0.143	0.110	0.273
AG	0.0	0.308	0.358

Table 2: Comparison of the coefficient of variation of the CPM values for the three different countries. Significant variance is observed, especially for the CPM maximum values.

spective country (e.g., the default gender parameter targets both male and female users).

Skewed distribution Our first observation is that the suggested bid data is highly skewed, with the median of the suggested CPM almost always being significantly closer to the minimum than the maximum. For example, if we target users in the United States, the minimum of the suggested CPM is typically between $0.03 and $0.07, the median is typically between $0.08 and $0.18, and the maximum is typically between $0.80 and $2.00. This property holds regardless of the targeting parameters that we choose.

Significant variance Our next observation is that multiple suggested bids with the same targeting parameters show significant variance over short time periods. For example, consider the graphs presented in Figure 3, which shows the 1,000 suggested CPM bids for three different countries with very different populations (United States, 159M users; New Zealand, 2.2M users; and Antigua and Barbuda, 29K users). In all three cases, the minimum, median, and maximum values show significant variance,[10] even from query to query (queries were roughly spaced 35 milliseconds apart). To quantify the variance, we calculate the *coefficient of variation* (the standard deviation divided by the mean) of the distribution, and present the results in Table 2. In almost all cases, significant variance is observed, with the CPM maximum always showing the highest coefficient of variation.

Variance independent of audience size Our third observation is that the variance observed is independent of the audience size. We compare the audience size versus the coefficient of variation of the CPM minimum, median, and maximum values for each country in Figure 4. Across suggested bids from the 204 countries, we observe no correlation between the audience size and the coefficient of variation of any of the CPM values: the correlation coefficients are -0.02 (CPM minimum), -0.08 (CPM median), and -0.03 (CPM maximum). We observe similar results with audiences derived from over 100 different sets of targeting parameters including US states and zip codes, user interests, and relationship status.

Variance across accounts Our fourth observation is that the suggested bids queried at the same time from different accounts show no correlation. To explore this, we used our multiple Facebook accounts described above and queried for suggested bids for the same target demographic at the same time from multiple accounts. Despite using a range of tar-

---

[10]The minimum and median values for Antigua and Barbuda take on fewer values, but this is likely an artifact of the $0.01-granularity of the data returned.

(a) United States (159,115,060 users) (b) New Zealand (2,272,620 users) (c) Antigua and Barbuda (29,580 users)
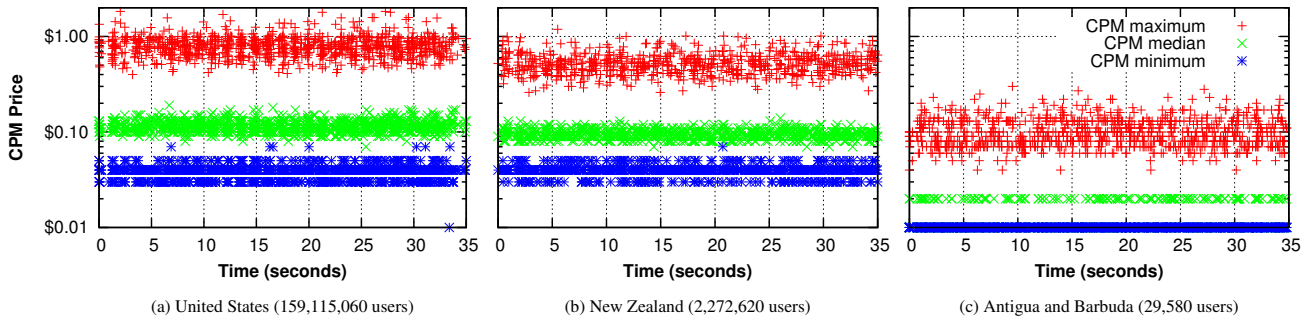
Figure 3: 1,000 suggested bids for three different sets of targeting parameters with massively different audience sizes (note the log-scale on the *y*-axis). Significant variance is observed, as well as a skewed distribution, across all three countries.

geting parameters (resulting in a variety of audience sizes), we did not find any correlation between the suggested bids received by the different accounts (despite the fact that the queries were issued at the same time).

An example is shown in Figure 5, containing the CPM maximum value received by two accounts targeting US users. Visually, we can observe little correlation between the values returned by the two accounts; for example, the spike around 85 seconds observed by account 2 is never reflected in account 1's results. Moreover, the Pearson's correlation coefficient between the values received by the two accounts is -0.01, further indicating the lack of a correlation.

Non-persistence of min or max Our final observation is that the CPM minimum and maximum values do not "persist" from query to query. For example, consider the graph in Figure 3 (c), which shows the CPM data for Antigua and Barbuda. While there are only 29,580 Facebook users in the country (as measured by the audience size), the CPM maximum value varies repeatedly between $0.04 and $0.30 in less than 100 milliseconds. This indicates that the minimum and maximum are very likely not calculated from the same pool (or a rolling pool) each time; instead, as we describe in Section 3.4, we believe they are calculated by *sampling* from the pool of recent winning bids.

3.3 Reverse-engineering suggested bids

The suggested bid feature is not documented by Facebook; the most relevant documentation describes the purpose of

the feature as helping advertisers select a bid that is likely to cause their ads to be shown to users [13]. We requested additional information from Facebook's Advertising Support Team on how the suggested bids are calculated and received the following information (emphasis ours):

> The suggested bid range you see when creating your ads is based on the bids that are *currently winning the ad auction* for the users you've chosen to target.

Thus, it is clear that the suggested bids are derived from the *recent winning bids* on the target users, but it is not quite clear *exactly how* they are derived.

Ultimately, the suggested bid algorithm is a black box; we are unlikely to be able to definitively reverse-engineer how they are calculated. Instead, we look for the most reasonable explanation for how suggested bids are derived given our observations. We present three hypotheses below and rule out two as unlikely.

Hypothesis 1: Winning bids change rapidly The first hypothesis is that the suggested bids are derived from the most-recent-k winning bids for the target users (for some value of k). If this were the case, the observed variance would be due to the set of recently-won bids changing rapidly. However, this hypothesis does not explain the significant variance observed on very short timescales for countries with very small audience sizes (e.g., the Antigua and Barbuda CPM from Figure 3 (c)); with such small audiences, it is unlikely that ads are served to these users quickly
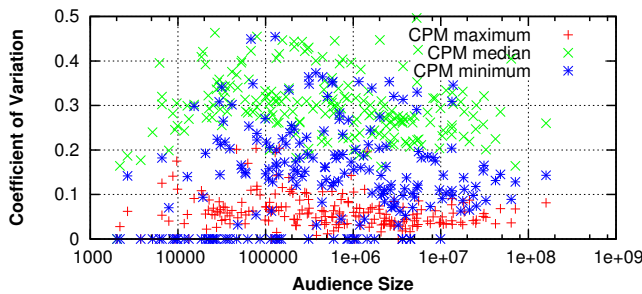
Figure 4: Audience size vs. coefficient of variation (standard deviation divided by mean) of suggested CPM bids for all 204 countries. No correlation is observed for CPM minimum, median, or maximum.
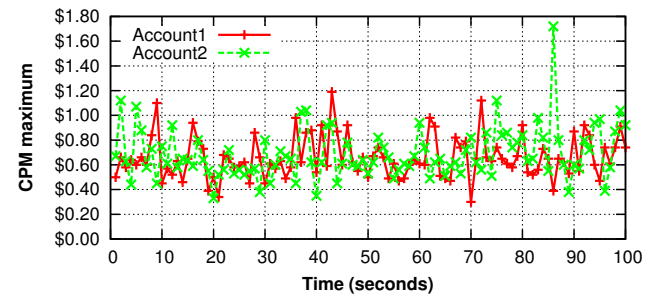
Figure 5: CPM maximum values for 100 successive suggested bids targeting US users, queried from 2 different accounts at the same time. The prices received by the two different accounts show no correlation.
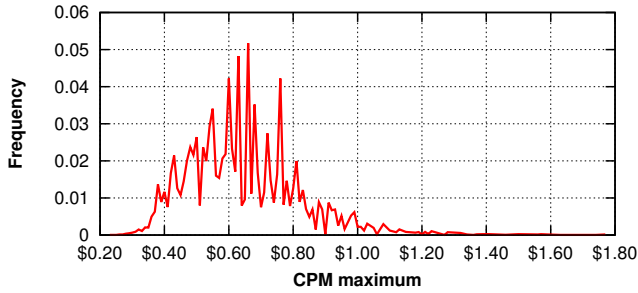
Figure 6: Example probability distribution function of CPM maximum values for 20,000 successive suggested bids (US 25-year-old females interested in computer programming; 179,760 users). This distribution fails statistical tests for multiple common distributions, suggesting the absence of random noise.

Figure 8: Diagram of Hypothesis 3: Sampling winning bids. To generate each suggested bid, Facebook selects a sample of bids from the set of all bids, and returns the median, minimum, and maximum from the sample.

enough to account for the rapidly changing minimum and maximum values.

Hypothesis 2: Adding random noise The second hypothesis is that Facebook is adding random noise to the returned suggested bid data (possibly to obfuscate the true value). To explore this hypothesis, we collected 20,000 suggested bids in quick succession for a small target population (25-year-old U.S. females interested in computer programming; 179K users); we then ran a number of statistical tests to see if the data matched a number of common statistical distributions (as would be expected were Facebook adding random noise): Uniform random, Gaussian, Cauchy, Log-Normal, or Logistic distributions. We found a poor fit for all distributions, with a p-value of less than 10^{-16}. An example probability distribution function of one of these suggested bid sets is shown in Figure 6.
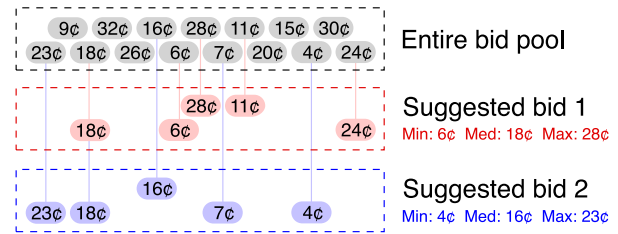
Hypothesis 3: Sampling winning bids The third hypothesis is that Facebook is *sampling* from the recent-k winning bids, and is reporting the minimum, median, and maximum of the sample; a diagram of this process is presented in Figure 8. This hypothesis is consistent with the format of the returned data (i.e., the fact that Facebook returns a minimum, median, and maximum) and is also consistent with our observations: we would expect to see significant variance from query to query, as different samples of the recent winning bids are used to generate each response. Moreover, this hypothesis is compatible with the variance being independent of audience size if the same k is used.

While verifying the correctness of Hypothesis 3 is likely only possible if one has access to Facebook's internal systems, we rely on the fact that it is both a logical mechanism for calculating suggested bids and is consistent with all of our observations on the properties of suggested bids.

3.4 Validating suggested bids

We now validate that suggested bids can be used to measure the overall Facebook ad market by showing that suggested bids reflect changes to the marketplace and that they correlate with Facebook's revenue.

Changes to the market To test our hypothesis that suggested bids are generated using recent winning bid data, we

ran an experiment where we actively participate in the advertising market to see how quickly changes propagate into the suggested bids. To do so, we chose a small country (Seychelles, 26K Facebook users) with a low suggested CPM. We then created three accounts[11] to advertise with, and submitted one ad campaign from each targeting Seychelles users. To visibly affect the market, each of our three advertising accounts bid a significantly higher CPM ($1.00) than the suggested CPM maximum ($0.16). We ran the advertising campaigns concurrently for 8 hours, receiving an average of 19,903 impressions to 3,543 users.

To measure the effect the campaigns had on suggested bids for Seychelles users, we also collected data on the suggested bids for targeting Seychelles users every 5 minutes using a separate account. In order to observe the changes, we started collecting this data 8 hours *before* all three of the advertisement campaigns became active; we also collected suggested bid data for the 8 hours during the campaign and for 44 hours *after* the campaigns ended.

Figure 7 presents the results[12] of this experiment, showing the CPM medians and maximums before, during, and after our campaigns (the campaign was running during the shaded region). From Figure 7 (a), we observe that the CPM median value was stable at $0.01 before we started our campaigns, rose up to $0.07 within 18 hours after our campaigns, before returning back to $0.01. Examining Figure 7 (b), we find that the suggested bids maximum rose dramatically from $0.16 up to $7.64; after we paused our campaigns, it fell back to a low price $1.47. Overall, this experiment shows that changes to the ad market are reflected in the suggested bids, and provides evidence for our hypothesis that the suggested bids data comes from a sample of the recent winning bids.

Comparison with Facebook's revenue Finally, we compare the data we observe via suggested bids to the only ground-truth that we know of: Facebook's SEC filings. In Facebook's March 2013 10-Q filing, Facebook reports the Average Revenue Per User (ARPU) at the granularity of region (US+Canada, Europe, Asia, and the Rest of the World). To compare the suggested bid data to Facebook's ARPU, we aggregate our CPM median data into the same regions and take the average across all countries in the region

---

[11] We choose to create three accounts, rather than one, as Facebook places multiple advertisements from different advertisers on each page.

[12] The fractional values result from a change by Facebook on May 1, 2013 to provide more precise suggested bids.
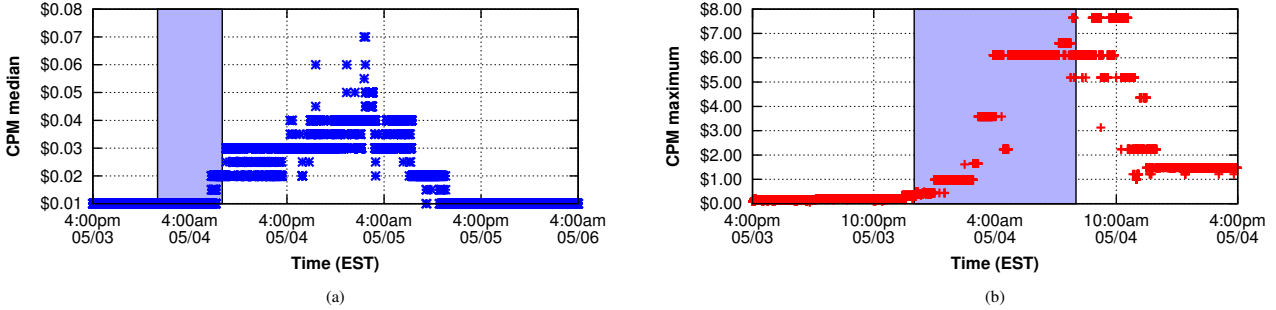
Figure 7: (a) CPM median values and (b) CPM maximum values before, during, and after our three advertising campaigns targeting Seychelles users (the shaded region represents the time of the ad campaigns. We observed that our advertising campaigns were quickly reflected in the suggested bids.

(weighted by audience size). We then scale both Facebook's ARPU and our aggregated suggested bid data relative to the US+Canada region. Of course, aggregating suggested bids in this way ignores many aspects of how revenue is generated (e.g., the activity level of different demographics), but can provide rough guidance on the relative revenue for different regions.

The results of this experiment are presented in Table 3. We observe similar trends between the two measures: Both Facebook's ARPU and the suggested bids rank the regions in the same order, with the Europe and Rest of the World regions at approximately the same ratios. While our results are far from being conclusive, this result indicates that the suggested bid data that we obtain from Facebook's advertising pages at least correlates with the distribution of Facebook's revenue.

3.5 Using suggested bids

From the previous section, we conclude that the suggested bid data is most likely calculated by sampling from the recent winning bids for the target users. Since the properties of the suggested bids are somewhat unique, we now explore how researchers can use suggested bids to measure the Facebook ad market.

Multiple samples Given that each suggested bid is most likely generated from a sample of the recent winning bids, it is clear that a single suggested bid may misrepresent the overall bid distribution. Instead, we collate multiple samples together, extracting the overall minimum, median, and maximum from the collated samples (i.e., for the remainder of the paper, all reported minima are the minimum across

| Region | Facebook ARPU | | Suggested Bid |
	Raw	Scaled	Scaled
US, Canada	$3.50	1.00	1.00
Europe	$1.60	0.45	0.42
Asia	$0.64	0.18	0.30
Rest of World	$0.50	0.14	0.15

Table 3: Comparison of Facebook's ARPU and CPM median suggested bids. We scale all values relative to the US+Canada region. We observe the same ranking of regions.

multiple suggested bid minima; the same holds for median and maximum).

Convergence The next step is to choose *how many* samples to collate together. To do so, we examine how quickly different numbers of collated suggested bids converge towards the overall "true" minimum, median, and maximum. To explore this question, we use the 204-countries data from the previous section. Since we do not know the true distribution of the recent winning bids, we instead use the overall minimum, median, and maximum of each country's 1,000 samples in its place.

Figure 9 presents the convergence of each of the 204 countries towards its overall minimum (bottom), median (mid-
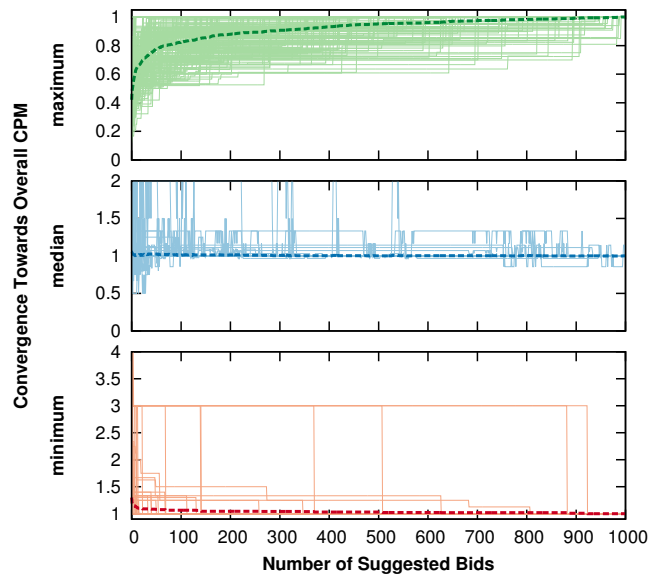
Figure 9: Convergence of different numbers of collated suggested bids towards overall CPM minimum, median, and maximum values for each of the 204 countries (each country is represented by a line in each graph). The average value across all countries is shown as the dark dashed line. Very quick convergence is observed for minimum and median, as expected.

dle), and maximum (top) values. Each country is represented by a thin line, and the average across all countries is shown as the dark, dashed line.

We observe that the minimum and median converge quite quickly: after only 25 suggested bids, both the minimum and median are within 15% of their eventual value (on average). Second, the maximum value converges more slowly (after 25 suggested bids, the maximum is within 30% of its eventual value, on average), which is expected due to the high variance and skewed distribution.

Choosing the number of suggested bids to collate together represents a tradeoff between accuracy and the load we place on Facebook. For the remainder of the paper, all of our reported data is the result of 25 collated suggested bids.

3.6 Limitations

Before examining the results for suggested bids for different user demographics, we first discuss a few of the limitations of our methodology.

User accounts As mentioned before, the advertising history of accounts could be a factor in determining the suggested prices (i.e., Facebook may show different suggested bids to different advertisers). Unfortunately, it is difficult (and expensive) to build an account with successful advertising history, as a result, to control any impact of this factor, all the user accounts used in our experiment are new created with no prior advertising history or uploaded content. Thus, the results in this paper are all comparable to each other, and we leave an exploration of the effect of advertising history on suggested bids for future work.

Facebook changes Our observations and methodology in this paper are based on the current advertising model used by Facebook; if Facebook makes internal changes in their advertising model, the suggested bid model, or the way they utilize the user data, our results may no longer be valid. However, we believe that our work provides researchers with a new approach for determining how suggested bids are calculated by Facebook, as well as ground-truth data to compare against.

Correlation versus causation In our analysis in the next section, we examine how the suggested bids are correlated with different user demographics. Of course, correlation does not imply causation, and it is possible that other, unknown factors are responsible for our observed correlations. Regardless, our analysis presents the first measurements of the relative value of different user demographics in OSNs.

4. ANALYSIS OF BID DATA

We now explore the properties of Facebook's ad auctions, using the suggested bid data.

4.1 Location

We examine how the *location* of the target demographic influences the ad auction winning bids, using the data set on 204 countries. We examine how the ad market CPM prices (represented by CPM median) correlate with the relative wealth of countries. To quantify the latter, we use GDP per capita [7], which is widely used in economics literature. The results are presented in Figure 10. As expected, we observe a correlation of 0.37 (statistically significant at the 0.001 level) between the GDP per capita and CPM prices. One
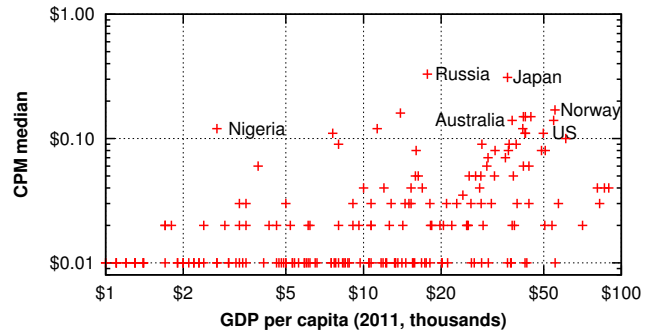

Figure 10: Scatterplot of CPM median versus GDP per capita for the 204 countries Facebook supports. Labeled are some of the countries with high CPM median values. All but a few countries have ad markets with very low CPMs.

notable outlier is Nigeria, which shows a CPM maximum on par with the U.S. while have a GDP per capita over an order of magnitude lower.

We now dig deeper into a single country and explore the differences in CPM prices between multiple cities in the same country. We choose to focus on the U.S., as it is the most mature Facebook ad market with the largest number of users. We query for the CPM prices of all U.S. cities with a population over 100,000 (285 cities). The results are presented in graphical form in Figure 11, plotting both CPM maximum (more red color representing higher values) and population (circle size) for each city. We observe that certain cities such as Las Vegas, NV and Hartford, CT show CPM maximums significantly higher than other cities like Cambridge, MA and Ann Arbor, MI, suggesting that certain cities have much more desirable users for advertisers to target.

Overall, we see dramatic differences in ad auction prices across different locations, with the most dramatic differences coming between users in different countries. As Facebook's ad markets continue to evolve, we can use our suggested bid methodology to measure their relatively maturity.
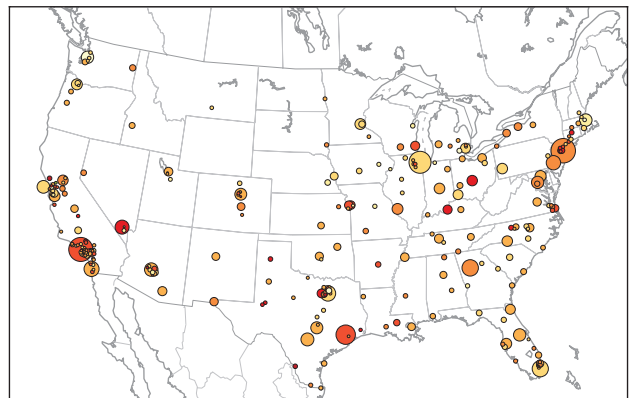
Figure 11: CPM maximum prices for all U.S. cities with population over 100,000. The color of each city corresponds to CPM maximum price (yellow to red represents increasing prices, and the size of the circle is in proportion to the number of Facebook users in each city.
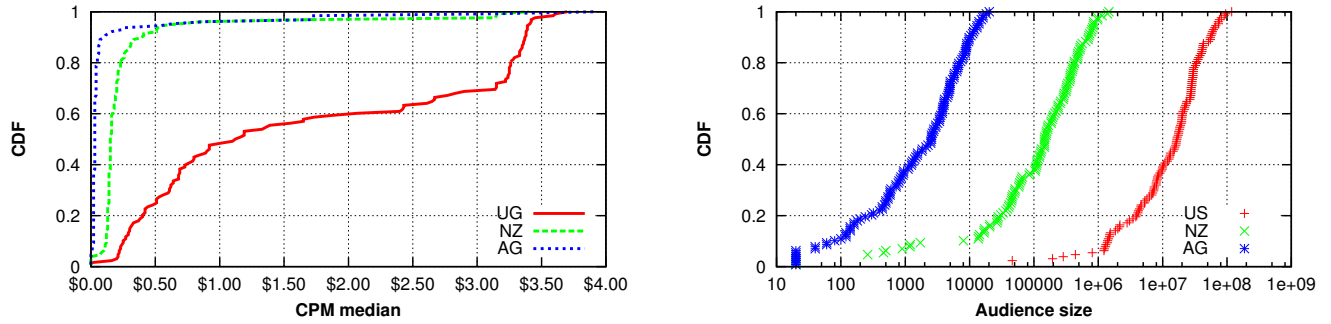
Figure 12: (a) Cumulative distribution of CPM median values for 129 interest categories for three countries. (b) Cumulative distribution of audience size for three countries (United States, New Zealand, and Antigua and Barbuda).

4.2 Age

We next explore how CPM median price is correlated with user *age*. We select the same three countries as before (the U.S., New Zealand, and Antigua and Barbuda), and retrieve suggested bids for users with different ages in each country. We note that Facebook's age policies come into effect here: the smallest age that an advertiser can target is 13, and targeting age 65 (the largest age one can target) encompasses all users 65 and over. The results of this experiment are presented in Figure 13 (top), showing the CPM median for different ages. We observe that in both the U.S. and New Zealand, as the age of target users increases, the CPM median price increases as well. The trend is less clear for Antigua and Barbuda, which we suspect is due to the smaller user population and less-well-developed ad market.

From Figure 13 (bottom), we observe that in all three countries, there is a rapid rise in the audience size (the number of Facebook users) between 13 to 18, followed by
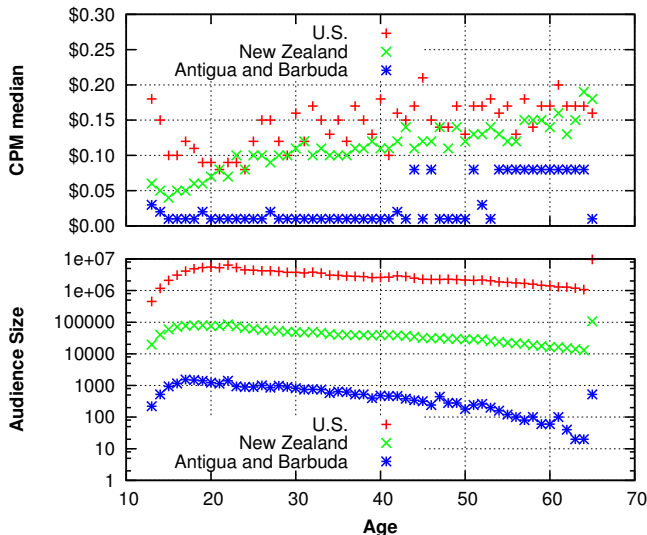
a slow decline over the remaining ages.[13] This observation corresponds strongly with previous studies about the age distribution of users in OSNs [35, 38].

4.3 Interests

For our final user attribute, we examine user *interests* in order to determine whether there are segments of the user population—based on particular interests—that are highly valued by advertisers. We are inspired by the wealth of studies on Web search auctions [26, 27, 31], which show there are certain keywords (e.g., "mesothelioma") that command prices thousands of times more than the average. For this experiment, we focus on *broad interests*, which are interests derived by Facebook based on user activity.
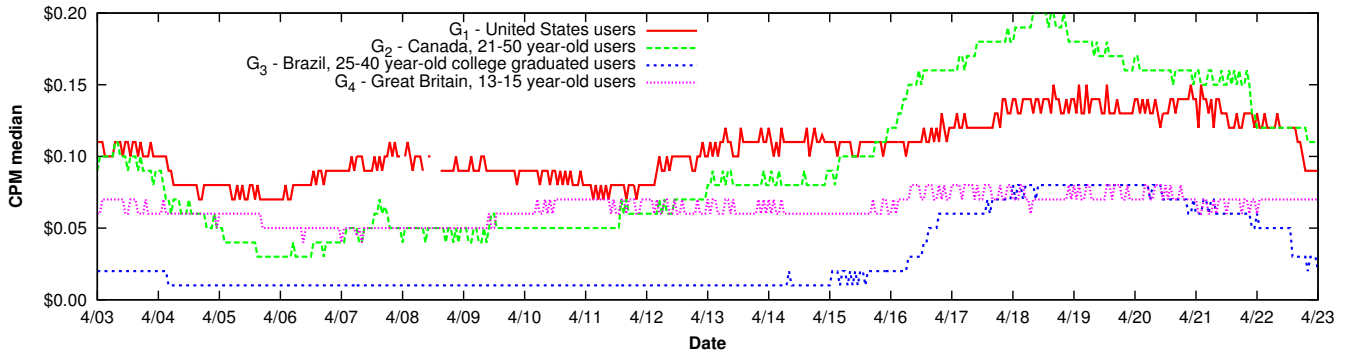
We retrieve suggested bids on all 129 Facebook-provided broad interest categories in each of the three countries that we have considered so far, and present the cumulative distribution across categories in Figure 12 (a). Surprisingly, the distribution of CPM median prices are rather broad: the most-expensive category is more than 20 times more expensive than the least-expensive category in all three countries. Interestingly, the most expensive categories in all three countries correspond to users who are traveling, recently engaged, or who like Apple products.

In Figure 12 (b), we observe that the distribution of audience size across different broad categories are widespread for all the three countries (note the log-scale on the *x*-axis). There are a significant number of broad categories are very popular within each country, for example, the largest categories in the United States are Mobile Users (All), Travelers, and Music (All).

4.4 Price stability

In addition to studying the CPM prices of different demographics, we also study the stability of different market prices over time. To do so, we select four different sets of targeting parameters, designed to cover a variety of targeting parameter types and audience sizes: G_1: U.S. users (167M users), G_2: 21-50 year-old Canadian users (11M users), G_3: 25-40 year-old college graduated Brazilian users (4.7M users), and G_4: 13-15 year-old British users (1.2M users). We track each of these sets of targeting parameters by retrieving 25 suggested bids each hour for a period of three weeks (April 3rd, 2013 through April 23rd, 2013).

Figure 13: (a) CPM median versus Age (b) Audience Size versus Age for three countries. CPM prices increase significantly with age for the United States and New Zealand, and massively different user populations are observed (note the log scale on the *y*-axis of the lower graph).

---

[13]The "jump" in audience size at age 65 is because 65 represents "65 and older."

Figure 14: Long-term tracking of CPM median prices over three weeks for four different sets of targeting parameters that show different properties with distinct trends over time.

Figure 14 plots the CPM median values for these groups during our period of observation. We observe that while there is significant short-term variance in many of the groups, there are also a number of longer-term trends present. For example, group G_1 shows a periodic increases spaced out approximately one week apart, and groups G_2 and G_3 shows a multi-day increase starting approximately on 04/16. Moreover, certain groups—such as groups G_2 and G_3—show significant fluctuations in price (up to six-fold), while others—such as G_4—do not vary much over the study period. Overall, our results suggest that there are significant short-term and long-term dynamics present in Facebook's ad auctions, which may be explained by the relative immaturity of different sectors of the market.

5. CONCLUSION

Advertising is now ubiquitous on the Web; most of the popular OSNs are funded via advertising on their site. While OSNs themselves have all information of their advertising models as well as how they utilize all the user data, there is little information that is shared with external researchers, advertisers, even the users themselves. Thus, OSN advertising markets remain quite difficult for researchers to study, even as these markets are growing in prominence.

In this paper, we have explored Facebook ad auctions through the *suggested bid* feature, showing how this feature can be used by researchers to make inferences on properties of the ad market. While we have only presented a subset of our experiments here, we have explored suggested bids for many other targeting parameters that Facebook provides, including gender, precise interests, relationship status, education, and workplaces. In all cases, we observed fewer differences than for the targeting parameters presented.

While we are far from covering the entire space—the set of possible targeting parameters is prohibitively large—our results suggest that advertiser interest is focused on location, user interest, and age parameters (consistent with ads in other media). Though our results are not necessarily surprising, we present the first mechanism for quantifying the relative value of different user demographics on today's OSNs. For advertisers, our work offers some guidance especially when they start to advertise on Facebook users; for users themselves, we provide them a basic idea about how valuable they are towards advertisers and Facebook, based on their own demographic information and activities which fall under different interests.

As the Facebook ad market continues to mature, we plan to repeat our analysis to study the evolution of the market. Moreover, as more OSNs develop advertising markets, our approach can be used to measure these markets as well; it is typical for ad markets to provide suggested bids. In other words, we can apply the proposed methodology to analyze the underlying models, and compare user value. For example, Facebook recently launched a real-time bidding ad system Facebook Exchange [12]; we aim to extend our methodology to measure this network as well. Additionally, both LinkedIn and Twitter have recently developed and deployed advertising markets; we hope to be able to apply our methodology to those sites.

Finally, we make all of our suggested bid collection code and collected data available to the research community at

`http://osn-ads.ccs.neu.edu`

6. REFERENCES

[1] S. Aral and D. Walker. Identifying Social Influence in Networks Using Randomized Experiments. *IEEE IS*, 26(5):91–96, 2011.

[2] L. Bin, S. Anmol, W. Udi, C. Jaideep, and G. Ramesh. AdReveal: Improving Transparency into Online Targeted Advertising. *HotNets*, College Park, MD, Nov. 2013.

[3] M. Bateni, J. Feldman, V. Mirrokni, and S. C.-w. Wong. Multiplicative Bidding in Online Advertising. *EC*, Palo Alto, CA, June 2014.

[4] Y. Bachrach, S. Ceppi, I. A. Kash, P. Key, and D. Kurokawa. Optimising Trade-offs Among Stakeholders in Ad Auctions. *EC*, Palo Alto, CA, June 2014.

[5] B. Cowgill and E. Zitzewitz. Corporate Prediction Markets: Evidence from Google, Ford, and Firm X. *EC*, Palo Alto, CA, June 2014.

[6] H. Cheng and E. C. Paz. Personalized click prediction in sponsored search. *WSDM*, New York, NY, Feb. 2010.

[7] CIA the World Facebook. `https://www.cia.gov/library/publications/the-world-factbook/index.html`.

[8] B. Eytan, E. Dean, Y. Rong, and R. Itamar. Social influence in social advertising: evidence from field experiments. *EC*, Valencia, Spain, June 2012.

[9] B. Edelman, M. Ostrovsky, M. Schwarz, T. D. Fudenberg, L. Kaplow, R. Lee, P. Milgrom, M. Niederle, and A. Pakes. Internet Advertising and the Generalized Second Price Auction: Selling Billions of Dollars Worth of Keywords. *Am. Econ. Rev.*, 97(1):242–259, 2007.

[10] J. Edwards. Facebook Shares Surge on First Ever $1 Billion Mobile Ad Revenue Quarter. 2013. `http://www.businessinsider.com/facebook-q4-2013-earnings-2014-1`.

[11] P. Furcy and K. Peter. Stochastic Variability in Sponsored Search Auctions: Observations and Models. *EC*, San Jose, CA, June 2011.

[12] Facebook Exchange: A New Way for Advertisers to Target Specific Users with Real-Time Bid Ads. `http://techcrunch.com/2012/06/13/facebook-exchange/`.

[13] Facebook Help: What is the Suggested Bid Range and how does Facebook calculate it? `https://www.facebook.com/help/213140778716849`.

[14] Facebook SEC Filings. `http://investor.fb.com/sec.cfm/`.

[15] A. Ghose and S. Yang. An empirical analysis of sponsored search performance in search engine advertising. *WSDM*, Stanford, CA, Feb. 2008.

[16] A. Ghose and S. Yang. An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets. *Manage. Sci.*, 55(10):1605–1622, 2009.

[17] A. Goldfarb and C. Tucker. Search Engine Advertising: Channel Substitution When Pricing Ads to Context. *Manage. Sci.*, 57(3):458–470, 2011.

[18] P. Gill, V. Erramilli, A. Chaintreau, B. Krishnamurthy, K. Papagiannaki, and P. Rodriguez. Follow the Money: Understanding Economics of Online Aggregation and Advertising. *IMC*, Barcelona, Spain, Oct. 2013.

[19] D. G. Goldstein, R. P. McAfee, and S. Suri. The Effects of Exposure Time on Memory of Display Advertisements. *EC*, San Jose, CA, June 2011.

[20] D. G. Goldstein, R. P. McAfee, and S. Suri. Improving the Effectiveness of Time-based Display Advertising. *EC*, Valencia, Spain, June 2012.

[21] Google. 2013 Financial Tables. 2013. `http://investor.google.com/financial/2013/tables.html`.

[22] Google AdWords. `https://adwords.google.com/`.

[23] Traffic Estimator. *Google AdWords*. `https://adwords.google.com/o/TrafficEstimator`.

[24] R. D. Hof. Advertisers Flock to Social Networks. *MIT Technology Review*, 2011. `http://www.technologyreview.com/article/424409/advertisers-flock-to-social-networks/`.

[25] J. Kim. adby.me. 2011. `http://adby.me`.

[26] L. Kim. The Most Expensive Keywords in Google AdWords. 2011. `http://www.wordstream.com/blog/ws/2011/07/18/most-expensive-google-adwords-keywords/`.

[27] Kidgas. Twenty Keywords with CPC More Than $20. 2011. `http://kidgas.hubpages.com/hub/Twenty-Keywords-with-CPC-More-Than-20`.

[28] S. Muthukrishnan, M. Pál, and Z. Svitkina. Stochastic Models for Budget Optimization in Search-Based Advertising. *Algorithmica*, 58(4):1022–1044, 2010.

[29] G. Noti, N. Nisan, and I. Yaniv. An Experimental Evaluation of Bidders' Behavior in Ad Auctions. *WWW*, Seoul, Korea, Apr. 2014.

[30] M. Ostrovsky and M. Schwarz. Reserve Prices in Internet Advertising Auctions: A Field Experiment. *EC*, San Jose, CA, June 2011.

[31] SpyFu. Keywords with Highest Cost Per Click (CPC). 2014. `http://www.spyfu.com/TopList.aspx?listId=3`.

[32] C. Tucker. Social Advertising. *SSRN*, Feb. 2012.

[33] D. Turnbull and L. F. Bright. Advertising academia with sponsored search: an exploratory study examining the effectiveness of Google AdWords at the local and global level. *IJEB*, 6(2):149–171, 2008.

[34] D. R. M. Thompson and K. Leyton-Brown. Revenue Optimization in the Generalized Second-price Auction. *EC*, Philadelphia, PA, June 2013.

[35] US Social Network Users by Age Group - New Statistics from Pew. `http://bit.ly/1tGTpso`.

[36] W. Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *J. Finance*, 16(1):8–37, 1961.

[37] H. R. Varian. Online Ad Auctions. *Am. Econ. Rev.*, 99(2):430–434, 2009.

[38] S. Wolfram. Data Science of the Facebook World. Apr. 2013. `http://blog.stephenwolfram.com/2013/04/data-science-of-the-facebook-world/`.

[39] S. Yao and C. F. Mela. A Dynamic Model of Sponsored Search Advertising. *Marketing Science*, 30(3):447–468, 2011.