From Social Trends to Profit: An Optimisation Model for Search Engine Adwords Management

Dong Wang[*§], Zhenyu Li[*], Mohamed-Ali Kaafar[†], Kavé Salamatian[‡], Gaogang Xie[*]
[*]Institute of Computing Technology, Chinese Academy of Sciences, China
[§]State Key Laboratory of Networking and Switching Technology, China
[†] NICTA, Australia, [‡] Université de Savoie, France
{wangdong01, zyli, xie}@ict.ac.cn, dali.kaafar@nicta.com.au, kave.salamatian@univ-savoie.fr

***Abstract*—In Search Engine Marketing (SEM), "third-party" partners play an important intermediate role by bridging the gap between search engines and advertisers in order to optimise advertisers' campaigns in exchange of a service fee. In this paper, we present an economic analysis of the market involving a third-party broker in Google AdWords and the broker's customers. We show that in order to optimise his profit, a third-party broker should minimise the weighted average Cost Per Click (CPC) of the portfolio of keywords attached to customer's ads while still satisfying the negotiated customer's demand. To help the broker build and manage such portfolio of keywords, we develop an optimisation framework inspired from the classical Markowitz portfolio management which integrates the customer's demand constraint and enables the broker to manage the tradeoff between return on investment and risk through a single risk aversion parameter. We then propose a method to augment the keywords portfolio with relevant keywords extracted from trending and popular topics on Twitter. Our evaluation shows that such a keywords-augmented strategy is very promising and enables the broker to achieve, on average, four folds larger return on investment than with a non-augmented strategy, while still maintaining the same level of risk.**

I. Introduction

Amongst the diverse forms of online marketing and advertising channels (*e.g.* email, mobile advertising), Search Engine Marketing (SEM), where ads are shown along with results of keyword queries, has been the fastest growing channel in the past decade. With up to \$19 billions, SEM revenues accounted for more than 38% of the total 2014 Internet advertising revenue [19].

In a typical SEM scenario, advertisers are allowed to publish their advertisements (ads in short) with the assistance of search engines in pages returned after search queries. The classic business interaction between advertisers and search engines involves the advertisers paying the search engines when their ads are being shown (Cost-per-Impression payment) or being clicked (Cost-per-Click payment). Search engines implement an online auction for deciding which ads to display in individual returned pages. Advertisers on the other hand bid on sets of keyword relevant to their business, called "ad keywords" or shortly as *"adwords"*, which will trigger the display of their ads in the returned search page for these adwords.

Generally, search engines provide to advertisers convenient platforms to manage their ads (such as Google AdWords, Bing Ads *etc.*). However, maintaining a profitable portfolio might still prove challenging for many advertisers, mainly due to lack of time, resources or expertise on ad markets. In this context, *"third-party" partners*, either advertising agencies (*e.g.* SuperMedia, Web.com), yellow page publishers or freelance consultants, play an important role of intermediaries between advertisers and search engine platforms. The interactions between third-party partners and advertisers create a *secondary market* in SEM (as opposed to the primary market where advertisers directly interact with search engines), where third-parties sell the services of optimising the advertisers campaigns while advertisers act as service buyers. We refer to third-party operators as "brokers" and advertisers as "customers".

In this paper, we present an economic model of the third-party market in SEM. Based on Google AdWords, a widely-used platform relying on the Cost-per-Click (CPC) payment mechanism, we first analyse the economic relation in the secondary Google AdWords market where customers and brokers negotiate their service costs. To this end, we model the relation with the truncated Constant Elasticity Demand model (CED). The analysis reveals that in order to optimise his profit while still being able to achieve the customer's demand, a third-party broker should minimise the weighted average CPC of the adwords portfolio.

Based on the CED model, we further develop an optimisation framework inspired from the classical Markowitz portfolio management which integrates the customer's demand constraint, and enables the broker to manage the tradeoff between Return On Investment (ROI) and the risk of his adwords portfolio through a single risk aversion parameter. This framework serves as a powerful tool for the broker as it illustrates well the efficient frontier: a curve which gives the optimal ROI for a given level of risk. The latter is useful for comparing different portfolio construction strategies and to make decision.

An intuitive strategy to maximise the return on investment is to select a set of adwords that have low CPC and high potential click numbers. In other words, the major challenge for a broker is then to build and manage such adwords portfolio. To this end, we consider Twitter as a possible source of "valuable" adwords. Our choice of relying on Twitter is motivated by recent research which reports on accurate predictions of the DJIA (Dow Jones Industrial Average) using elements of public mood inferred from Twitter [9]. We postulate that by referring to popular and trending topics on Twitter, a broker can foresee a set of adwords that are likely to attract high click numbers, while being not yet detected by other contestants (third-parties and advertisers) and have therefore low CPC. Indeed using adwords extracted from Twitter is not excluding adwords coming from traditional marketing method, *e.g.*, handpicking adwords by human marketing experience.

Using real-life data, we verify that trending and popular topics extracted from Twitter are plausible good candidates to feed the broker's optimal adwords portfolio. More importantly, we evaluate the application of our model and show that a broker could achieve a significant ROI improvement ($4\times$) over a classical portfolio management, while maintaining the same level of risk.

The rest of the paper is organised as follows. Section II analyses Google's adwords secondary market and proposes a portfolio optimisation framework. Section III explores strategies to build a portfolio of adwords. Section IV evaluates the proposed portfolio constitution method. Section V summarises related work. Finally, Section VI concludes the paper.

II. MODELING GOOGLE ADWORDS SECONDARY MARKET

In this section, we first analyse the Google AdWords secondary market and use a simple "Constant Elasticity Demand" based model to capture the broker's profit, and then propose a portfolio optimisation framework.

A. Broker's Profit Analysis

In our context, the valuable *product* that a broker provides to his customers consists of the adwords management and the optimisation service. In practice, the customer who wishes to maximise the impact of his advertisement campaign, entrusts a third-party broker to build an adwords portfolio. In return he pays the broker a service fee. The broker interacts directly with Google AdWords by choosing relevant adwords, setting the maximum bids while considering the overall budget, and pays the costs of displayed ads to Google. The broker's profit is simply the difference between the service fee paid by the customers and the advertising costs paid to Google.

Advertisers may have various aims for their advertisement campaigns, *e.g.* reaching a given audience in terms of click number, or achieving a given number of conversions, *i.e.*, clicks that result in other activities like buying the product or signing a petition, *etc.* The broker has therefore to align his profitability objective with the precise needs of his customer.

Typically, two types of contract between the advertiser and the broker are possible: (1) the advertiser has a target objective $D(a)$ for his ad a (a given click number or conversion number) at a time horizon $\mathcal{T}$ and the broker proposes an overall budget; (2) the advertiser has an overall budget for his campaign and the broker commits on the target objective $D(a)$ for this budget at time horizon $\mathcal{T}$. Regardless of the contract type, the efficiency of a third-party broker finally boils down to minimising the cost paid to Google.

Google AdWords uses a pay per click model, *i.e.*, the cost paid to Google AdWords depends on the Cost Per Click (CPC), which is determined through the online auction mechanisms and the click number. A reasonable fee strategy for a broker consists then of setting his service fee as function of the click number brought by the customer's ad, *i.e.* the broker sets a Price Per Click or Price Per Conversion (PPC) $P(a)$ for the ad a of his customer. The overall budget can be easily translated to PPC by dividing it to the target objective $D(a)$. As the cost paid to Google is measured in terms of click, for the ease of notation we assume that the target objective $D(a)$ is defined in term of clicks. We therefore consider that the contract between the broker and the customer is set on a Price Per Click (PPC)-basis, and that the customer's constraint (demand) in the contract is the number of clicks needed to achieve the objective in terms of clicks or conversions.

In order to satisfy the contract, the broker builds a portfolio of adwords, denoted $K(a)$ for the ad a. We first consider that the adwords portfolio $K(a)$ is *a priori* given in this section. We then detail the portfolio construction process in Section III. If, for each adword $i \in K(a)$ the $CPC_t(i)$ at time t, as defined by Google AdWords, and a PPC contracted with the advertiser $P(a)$, are given, then the profit of the broker up to time t from ad a can be expressed as the difference between his revenue and costs:

$$Q_t(a) = \sum_{i \in K(a)} \sum_{T(i,a) \leq t} (P(a) - CPC_{T(i,a)}(i)) \qquad (1)$$

where $T(i,a)$ is the set of time instants when the adword i was searched, the ad a was shown and a click was applied to the ad. While it is intuitive to consider that the high click number tends to result in a high CPC, Yuan *etc.* [30] have found that the bidding of advertisers is always unresponsive to the change of click number, meaning a stable CPC over time. This has also been confirmed by the analysis over our Twitter dataset (Section III-C).

We further use $S_t(i,a) = \sum \mathbb{1}_{T(i,a) \leq t}$ to represent the number of clicks on ad a resulting from adword i searches up to time t and $S_t(a) = \sum_{i \in K(a)} S_t(i,a)$ to be the total number of clicks on ad a up to time t. We can then simplify the expression of the broker's profit as:

$$Q_t(a) = \sum_{i \in K(a)} S_t(i,a) \left(P(a) - \overline{CPC_t(i)} \right) \qquad (2)$$

where $\overline{CPC_t(i)}$ is the average CPC of adword i up to time t. At time horizon $\mathcal{T}$ the broker will satisfy his contract if the committed demand in the contract is reached and his profit will become:

$$Q_{\mathcal{T}}(a) = \sum_{i \in K(a)} S_{\mathcal{T}}(i,a) \left(P(a) - \overline{CPC_{\mathcal{T}}(i)} \right) \qquad (3)$$

or equivalently

$$Q_{\mathcal{T}}(a) = D(a) \left(P(a) - \overline{\overline{CPC_{\mathcal{T}}(a)}} \right) \qquad (4)$$

where $\overline{\overline{CPC_{\mathcal{T}}(a)}} = \frac{\sum_{i\in K(a)}(S_{\mathcal{T}}(i,a)\overline{CPC_{\mathcal{T}}(i)})}{\sum_{i\in K(a)} S_{\mathcal{T}}(i,a)}$ is the weighted average CPC and the double bar notation indicates that the average is calculated both over time and over the adwords portfolio $K(a)$.

The Return On Investment (ROI) is therefore calculated as:

$$\overline{\overline{R(a)}} = \frac{\left(P(a) - \overline{\overline{CPC_{\mathcal{T}}(a)}}\right)}{\overline{\overline{CPC_{\mathcal{T}}(a)}}} \tag{5}$$

B. Demand modeling

Intuitively, a higher PPC indicates a higher revenue of the broker. However, the price has an immediate impact on the customer's demand. This relationship is generally characterised by a price/demand curve. We thus use a simple customer elasticity model, the Constant Elasticity Demand model (CED) [22]. The model assumes that the elasticity of the demand $\eta = \frac{\partial D(a)/D(a)}{\partial P(a)/P(a)}$ is constant, meaning that a relative increase (*resp.* decrease) in the price results in a proportional decrease (*resp.* increase) in the demand with a constant η. The CED model is widely used for describing the user utility on Internet [22]. In particular, it is appropriate for scenarios where the product demands are separable, *i.e.*, changes in demand or price for one product have no effect on others. These assumptions are valid for the Google AdWords secondary market where the demand $D(a)$ only depends on the overall budget and the PPC $P(a)$ on a but not on other ads.

In the CED model, the relationship between the customer's demand $D(a)$ for ad a and the PPC $P(a)$ is described by the following equation:

$$D(a) = \left(\frac{v(a)}{P(a)}\right)^{\alpha} \tag{6}$$

where $v(a) > 0$ is a valuation coefficient for a. The parameter $\alpha \geq 1$ is called *price sensitivity* and indicates the price elasticity of demand, *i.e.* $\eta = \frac{\partial D(a)/D(a)}{\partial P(a)/P(a)} = -\alpha$.

The unitary elastic case $\alpha = 1$ happens when the advertiser's budget is constant, *i.e.*, $P(a)D(a) = v(a) = cte$. However, to make the model more realistic, we should encompass the case where the customer sets an upper limit for the price $P_{\max}$ and a minimum number of expected clicks or conversions $D_{\min}$. The customer may decide to choose a different broker if the negotiated price is higher than $P_{\max}$ or if the broker cannot satisfy at least $D_{\min}$ objective of his ad. This suggests that a truncated CED model might be a better candidate to describe the customer's demand.

Adding the CED model into Eq. 4, the broker's profit becomes:

$$Q_{\mathcal{T}}(a) = \left(\frac{v(a)}{P(a)}\right)^{\alpha} \left(P(a) - \overline{\overline{CPC_{\mathcal{T}}(a)}}\right) \tag{7}$$

where $0 < P(a) \leq P_{max}$. The profit-maximising price $P^{\star}(a)$ for ad a can be obtained by solving $\frac{\partial Q_{\mathcal{T}(a)}}{\partial P(a)} = 0$ and considering the maximum price constraint. It can be expressed as:

$$P^{\star}(a) = \min\left\{\frac{\alpha}{\alpha-1}\overline{\overline{CPC_{\mathcal{T}}(a)}}, P_{\max}\right\} \tag{8}$$

Replacing $P(a)$ in Eq. 5 with $P^{\star}(a)$, we have the maximum ROI:

$$\overline{\overline{R^{\star}(a)}} = \min\{\frac{1}{\alpha-1}, \frac{P_{\max}}{\overline{\overline{CPC_{\mathcal{T}}(a)}}} - 1\} \tag{9}$$

The above analysis shows that the broker can optimise his profit by setting the selling PPC slightly higher with a coefficient $\frac{\alpha}{\alpha-1}$ than the weighted average CPC. However in practice, the Google AdWords market is very dynamic, which leads to the dynamics analysis of ROI.

C. Dynamics Analysis of ROI

The instantaneous value of the CPC for an adword, the number of clicks and the conversion rate are stochastic processes which vary with time. As such, ROI is a random variable with mean $\mathbb{E}\{\overline{\overline{R(a)}}\}$ and variance $\sigma^2(\overline{\overline{R(a)}})$. For an ad a attached with a portfolio of adwords $K(a)$, the average and variance of ROI can be derived as:

$$\begin{cases} \mathbb{E}\{\overline{\overline{R(a)}}\} & = \sum_{i\in K(a)} w_i R(i) \\ \sigma^2\left(\overline{\overline{R(a)}}\right) & = \sum_{i,j\in K(a)} w_i w_j \sigma\left(R(i)\right)\sigma\left(R(j)\right)\rho(i,j) \end{cases} \tag{10}$$

where $w_i = \frac{S_{\mathcal{T}}(i,a)}{D(a)}$ is the weight of adword i in the portfolio, *i.e.*, the proportion of clicks or conversions resulting from the adword $i \in K(a)$ among all clicks or conversions leading to the ad a satisfying the demand, $R(i)$ is the ROI of adword i and $\rho(i,j)$ is the correlation coefficient between the ROI of i and j.

Again, the overarching objective of a broker is to maximise his ROI by satisfying the customer's demand $D(a)$ on an ad a, while minimising the stochastic risk resulting from market fluctuations. In such a context, the risk for the broker is that the final $\overline{\overline{CPC_{\mathcal{T}}(a)}}$ becomes larger than $P(a)$ resulting in loss. In order to protect the broker from such a risk, we adopt an approach inspired from the Markowitz formulation of portfolio optimisation in financial market [21]. The Markowitz portfolio optimisation defines the proportion of capital that an investor should dedicate to different assets with different ROIs and risks, in order to maximise his profit given a level of risk aversion. In our approach, the stochastic risk of the broker is captured by $\sigma^2\left(\overline{\overline{R(a)}}\right)$ and his aversion to risk is characterised by a value $\gamma > 0$ named *risk aversion coefficient*. The larger γ is, the more risk the broker is ready to take in order to increase his ROI. In this case, the stochastic version of the broker optimisation problem can be written as:

$$\min_{\mathbf{w}\in\mathbf{\Delta}}\left(\sigma^2\left(\overline{\overline{R(a)}}\right) - \gamma\mathbb{E}\{\overline{\overline{R(a)}}\}\right) \tag{11}$$

where $\mathbf{w} = (w_i), i \in K(a)$ is the vector of weights and Δ is the simplex surface $\{\mathbf{w} \in [0,1]^{|K(a)|} | \sum_{i\in K(a)} w_i = 1\}$.

The major difference between the classical Markowitz formulation and our formulation comes from the constraint to achieve the customer demand $D(a)$ rather than only trying to maximise the portfolio ROI as in classical Markowitz. In other words, we use the Markovitz formulation with demand constraint case. Moreover, in classical Markowitz formulation the share of the capital assigned to each asset w_i is a deterministic value that is set at the time of the constitution of the portfolio, while in our formulation the w_i is a random variable depending on the willingness of search engine users to engage with the ad, and on the relevance of the ad content to the adword i (*i.e.* Quality Score [2][1]).This difference becomes more important when the customer's demand is described through a conversion number, because this adds one more element of randomness: the decision of the viewer to convert a click into a concrete action like buying the product.

We deal with the randomness of w_i in two ways. The first approach takes some cautions with the official click number estimates provided by Google AdWords, *e.g.*, limits the value of w_i to a percentage $0 < \beta \leq 1$ of the click number estimated during the optimisation, *i.e.*, $w_i < \frac{\beta S^g_{t_0}(i,a)}{D(a)}$ where $S^g_{t_0}(i,a)$ is the click number estimate given by Google AdWords at the time of decision t_0. However, this approach is not available for the conversion demand at the beginning of an ad campaign as the conversion rate of this ad is unknown. The second approach that is applicable when the target demand is a conversion number, is a dynamic version of the first approach, where the weight w_i for each adaptation period (*e.g.* each day) is limited by the number of clicks or conversion rate observed so far in the previous observation periods. The weights are unlimited during the first period where previous observations are missing and this entails a re-optimisation process at the beginning of each period. Note that by dealing with the randomness of w_i, we incorporate Quality Score into our optimisation model.

The above analysis leads to a useful concept of portfolio management: *efficient portfolio*. A portfolio is called "efficient" if it has the best possible expected ROI for its level of risk, $\sigma^2\left(\overline{\overline{R(a)}}\right)$. The efficient portfolio is illustrated through the risk/ROI plane, a frontier separating achievable risk/ROI tradeoffs (on the right of the curve) from the unachievable one (on the left of the curve), as shown in Fig. 1. The points on the efficient frontier can be calculated by solving the optimisation in Eq. 11 for different values of risk aversion γ and plotting the resulting optimal ROI and risk. In the Modern Portfolio Theory, this efficient frontier is always used as a metric to compare different portfolio constitution approaches. A top-left oriented frontier means that higher ROI with lower risk is achievable. We utilise this frontier later for evaluation.

III. Building Portfolio with Twitter

So far, we presented the theoretical foundations for optimising an adwords portfolio by a third-party broker, where the portfolio is assumed to be a priori given. However,

[1]Quality Score (QS) defined by Google ranges from 1 to 10 for each pair (query, ad) that measures the relevance of the ad content and the search query. Google weights the CPC bidden by the advertisers by the QS, and decides therefore the auction winner and the rank of the ad display by combining the CPC and the QS [2].

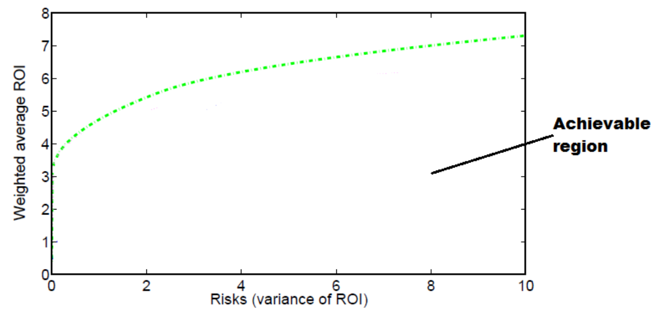


Fig. 1. An illustration of efficient frontier

building an adwords portfolio in practice is far from being trivial because of the dynamics of adwords market. In this section, we first provide an overview of possible ways to constitute a portfolio and then propose to augment portfolios with additional adwords from Twitter topics. Unless stated otherwise, we consider the customer's demand to be expressed in terms of number of clicks.

A. Overview of portfolio constitution

Google provides a "Keyword Planner" tool for helping users choose and match adwords [4] to their ad campaigns. Through this interface, users can select and test several combinations of adwords portfolio. Using "Keyword Planner", the aim of the broker would be to uncover the adwords with low CPC and high potential of clicks or conversions. However, such adwords are likely to attract competition quickly and their CPC are likely to increase fast in the future.

Generally, two approaches might be considered in this context. The first approach consists of searching for "long-tail" queries, *i.e.*, infrequent queries that are likely to draw targeted visitors on ads, *e.g.* the bulk of Amazon's revenues comes from a long tail of items but not from a few block-buster items [6]. Several business sites are targeting such keywords in search engines recently [24][7]. The second approach consists of exploring the adwords space for promising topics which have not yet attracted the interest of competitors (third-party brokers and advertisers), but have already generated a surge in search traffic and as such are likely to be efficient from a user interest perspective. This aims not at replacing the first approach but rather at augmenting it.

We consider the second approach as one possible strategy to build and augment the adwords portfolio. Previous researches revealed stock market changes can be predicted based on observations of Twitter trends [31][9][25][11]. This motivates us to consider topics originating on Twitter as potential candidate adwords for an efficient portfolio.

B. Collecting adwords from Twitter

Any word or sequence of words mentioned in tweets can be considered as a potential *topic*, i. The popularity of a topic i is defined as the number of tweets mentioning it. Topics mentioned relatively more frequently over a time period are called *popular topics*. *Trending topics* on the other hand are defined by Twitter as topics with a popularity that is increasing relatively faster than other topics.A "non popular" topic might be "trending" when its popularity increase rate is large, while a topic remaining popular for a long time is not likely to stay "trending", as the number of tweets mentioning this topic stabilises. We stratify the topics extracted from Twitter into three classes: *trending*, *popular* and *normal* (*i.e.* random). In what follows, we demonstrate our approach through data collected from Twitter.

Trending topics: We extracted the topics provided publicly by Twitter as trending over the period spanning from Oct. 26th to Nov. 1st 2013 and from Feb. 2nd to 8th 2014. Specifically, every five minutes during the crawling period, we collected the top 10 trending topics in US suggested by Twitter. A topic can be trending for more than one day, but we only considered the trending topics which had never been trending before the sampling day. This resulted into 1,175 unique trending topics that composed the trending topics dataset T. We use the two time periods to catch the trending topics related to candy (*e.g.* "Halloween") and sports (*e.g.*"super bowl" and "world series") respectively, which will be used for analysis and evaluation later.

Popular topics: We used the Twitter streaming API to crawl a set of tweets over the same periods as above. This resulted into 105,946 tweets randomly sampled by Twitter API. We binned all sampled tweets into subsets with daily granularity and then extracted for each daily subset S_i the word frequency for each word $w \in S_i$. Note that we ignored stop words (*e.g.* "a", "after", "that", *etc.*) which naturally appear with higher frequencies (readers could refer to [5] for a complete list of the stop words). A word was counted only once per tweet even if it was repeated in the tweet. We also filtered out hashtags consisting of more than three words as they are too long to make adwords. After the data preprocessing, we chose the top-200 most frequent words for each daily subset, leading to 2,800 popular topics out of 35,705 topics extracted over the two weeks. Again, we also removed the duplicate topics and only considered the topics which had never been categorised as popular before the sampling day. This resulted into 1,214 unique popular topics and constituted the dataset P. Although only 3.4% of potential words (topics) were chosen, these represented more than 30% of the total popularity (in terms of volume of tweets mentioning them) over all extracted words.

Normal topics: From the same set of sanitised tweets crawled for popular topics, we also randomly chose 1,000 other words and considered them as our normal (random) topics dataset N. This set will be used as a comparison with the two others.

For all 3,389 topics in the three Twitter datasets, we used the "Keyword Planner" tool of Google AdWords to collect the daily CPC and number of daily clicks estimates for the 10 days following the first day each topic was considered as trending/popular or randomly sampled from Twitter. As the metrics provided by Google slightly change over time due to Google's extraction procedure [23], we sampled for each day 8 time points (once every three hours) and then used the average value over these samples as a single daily metric for each adword. All CPC values are in US dollars. In order to reduce the randomness and uncertainty related to the auction, we set daily budget and max CPC bid as the maximum values allowed by Google so that the estimates returned by Google are the estimated maximum CPC and the estimated maximum click number. This means that the ROI we are obtaining are lower bounds, *i.e.*, the broker can hope to achieve higher ROI than what is reported in this paper.

TABLE I
GOOGLE ADWORDS PROPERTIES OF TOPICS IN THE TWITTER DATASETS

Dataset	Average	Variance	Median	1-prct	5-prct	95-prct	99-prct
$\overline{CPC(N)}$	3.52	22.71	2.15	0.001	0.02	11.88	24.63
$\overline{CPC(P)}$	3.54	19.47	2.30	0.004	0.05	10.82	20.59
$\overline{CPC(T)}$	3.42	19.40	2.23	0.004	0.08	12.93	25.42
$\sigma^2(CPC(N))$	1.56	4.07	0.86	0.004	0.02	7.58	16.26
$\sigma^2(CPC(P))$	1.73	7.96	0.80	0.008	0.04	7.32	14.95
$\sigma^2(CPC(T))$	1.86	8.95	0.87	0.008	0.08	6.33	9.97
$\overline{Clicks(N)}$	329.7	7.8×10^5	35.9	0.01	0.10	1866	5166
$\overline{Clicks(P)}$	928.9	1.5×10^7	51.3	0.03	0.25	4490	18332
$\overline{Clicks(T)}$	476.0	3.5×10^6	34.49	0.06	0.26	1934	7841
$\sigma^2(Clicks(N))$	109.0	4.8×10^4	23.3	0.03	0.0.17	554	1285
$\sigma^2(Clicks(P))$	224.5	6.4×10^5	25.1	0.07	0.29	904	4390
$\sigma^2(Clicks(T))$	162.1	3.2×10^5	20.91	0.07	0.31	547	3445

The words with at least one non-zero CPC value in the 10 days account for 50% of the topics in the normal topics dataset, for 68% in the popular topics dataset and for 63% of the trending topics. These percentages are significant and also show that our approach is applicable in practice for the adwords augmentation. The zero-CPC words are not considered in our study since they are inactive in Google AdWords and as such we are unable to evaluate the process of using them.

C. Analysis of the Twitter topics

We show in Table I the relevant statistics derived for topics that have at least a non-zero CPC in the 10 days of Google AdWords monitoring, *i.e.*, topics that are active in Google AdWords.

The statistics show that the daily average CPC and the corresponding variance of the CPC are very stable across the three datasets. A non parametric Kolmogorov-Smirnov distribution test could not reject the hypothesis that the three datasets come from the same distribution (using a 5% significance level). However both the distribution of average CPC and variance of CPC are highly skewed as the medians are far from the averages. The unbalance is mainly due to the tail, that is, some very large values pull the mean away from the median.

The daily number of clicks shows different distribution statistics across the three datasets. The estimate of daily number of clicks for popular topics is larger than the other two. Interestingly, the comparison of normal and trending topics shows that while the medians are close, the average estimate of the number of clicks of trending topics is significantly larger than the normal ones, indicating that there are more topics in the tail of trending topics with very large number of clicks. This suggests that with similar CPC values (prices), the broker can expect a larger average number of clicks for popular and trending topics. According to our analysis in Section II, the

TABLE II
CLICKS GROWTH FOR THE THREE DATASETS

Dataset	Average	Median	95-prct
Growth rate in N	5.14	1.19	64.74
Growth rate in P	10.04	1.16	182.45
Growth rate in T	11.37	1.29	252.42

$\overline{\overline{CPC_{\mathcal{T}}(a)}}$ controls the ROI of ad a, therefore a higher number of clicks for an adword with a stable CPC means a lower $\overline{\overline{CPC_{\mathcal{T}}(a)}}$ for the broker and a higher profit.

Lastly we observe that the variance of click number estimates shows a significant difference between the normal stratum and the other datasets. The variability in terms of click numbers for popular and trending topics is higher than normal topics. In order to evaluate whether this should be interpreted as a higher risk or a higher opportunity for popular and trending topics, we analyse the estimate of clicks increase. Specifically, for each adword, we fit the 10 daily click estimate values into a linear regression function and extract the growth rate of the estimate click number as the slope of the fitted function. We provide in Table II the relevant statistics which show that the higher variability is in fact an opportunity, as the number of clicks for popular and trending topics is growing faster than the normal topics. It is noteworthy that the medians for normal and popular topics are very close and again the tails of popular and trending topics are the key difference.

In summary, the analysis shows that while there is not a significant difference in the CPC of adwords originating from Twitter topics, there is a major benefit in terms of number of clicks to add the adwords extracted from popular and trending topics into the portfolio.

D. Portfolio constitution methodology

The great potential of the popular and trending topics on Twitter in improving the ads clicks motivates our portfolio constitution methodology. In detail, a broker follows two steps to build an efficient portfolio. At first, he generates an initial reference portfolio using either adwords suggested by Google (*e.g.* via "Keyword Planner"), or any of the numerous methods developed in the past couple of years for adwords portfolio selections [20][14][26], or even random adwords portfolio selections. In the second step, the broker looks at trending and popular topics coming from Twitter and augments his reference portfolio with relevant trending and popular topics. In other words, we are aiming not at replacing the existing methods, but rather at augmenting them with topics from Twitter. Finding relevant adwords amongst thousands of trending and popular topics may prove challenging. One possible approach is to use ontologies that can characterise the semantic proximity of keywords. Such ontologies can be built through human expertise or automatically using Wikipedia [12]. This will facilitate the search for relevant trending and popular topics.

It is noteworthy that our objective in this paper is not to evaluate the initial reference portfolio selection itself but rather to show a portfolio augmentation technique and to suggest the interest of adding adwords coming from Twitter popular and trending topics. As such we do not compare nor describe these advanced methods of adwords selection but rather simply use Google AdWords suggestions and random adwords portfolio selection.

TABLE III
ADWORDS USED IN OUR TWO SCENARIOS

Candy seller		Sports apparatus e-shop	
Google	Twitter	Google	Twitter
candy	*halloween*	*sports*	*nba*
online	*halloween images*	*ball*	*clipper*
chocolate	*halloween quotes*	*football*	*real madrid*
bar	*happy halloween*	*baseball*	*world series*
shop	*trick or treat*	*ride*	*super bowl*

Our methodology is best explained by two case studies: an online candy seller and a sports apparatus e-shop, both of which are assumed to contact a broker to start up Google AdWords campaigns. We use the "Keyword Planner" of Google for initial reference portfolio selection as in addition to price estimation of adwords, the "Keyword Planner" can also provide a set of suggested adwords for a specific product. We make use of these two functions to find relevant adwords for ads and build an initial reference portfolio. The fact that the initial reference portfolio is derived using "Keyword Planner" ensures that adwords in the reference portfolio have estimates (coming from Keyword Planner) for the average daily CPC, the variance of the CPC and the click numbers. Using these values, the broker first checks the range of click numbers for which the reference portfolio is feasible, *i.e.*, the set of click numbers which he can commit achieves his customer's demand. He thereafter derives the maximum average CPC with minimum risk $\overline{\overline{CPC^{+}(a)}}$. This latter value is obtained by deriving the optimum portfolio with risk aversion $\gamma = 0$. $\overline{\overline{CPC^{+}(a)}}$ is then used to set the price $P(a)$.

For each of these two scenarios we assume that the first day in our dataset is the decision day for the broker, and the broker generates an initial portfolio of adwords at the decision day. This initial portfolio is used for two purposes: to set the price $P(a)$ negotiated with the customer by broker and to be a reference portfolio compared with other strategies. Thereafter, the broker looks at trending and popular topics and chooses some of them to augment his portfolio.

Specially, to build the initial reference portfolio, the broker first queries "Keyword Planner" to get the top-5 suggested adwords as shown in Table III. We augment such reference portfolio with 5 relevant topics extracted from the trending and popular topics in our Twitter dataset. As our data gatherings happened in November and February, we have trending topics mentioning "Halloween" for the candy ad and mentioning "world series" and "super bowl" for the sports ad.

We evaluate the reference versus augmented portfolios for the above two scenarios in Section IV. Naturally, two particular scenarios are not enough to validate our approach. We thus also utilize the random adwords selections as the initial reference portfolio constitution method, a technique frequently used in stock market studies [10]. Again, it is noteworthy that our adwords augmentation using popular and trending topics on Twitter is independent of the initial portfolio constitution methods.

IV. Applications and Evaluation

In this section, we evaluate the adwords portfolio constitution method using our optmisation model. The aim is to evaluate whether the portfolio management methodology we developed is able to achieve high ROI with low risk.

A. Evaluation methodology

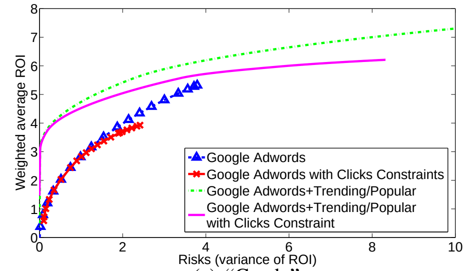
First, we present the challenges faced to make a meaningful evaluation of our research. We contacted two broker companies (Adobe, 4-traders) and these were not willing to provide details about their methodologies of selecting adwords as obviously these were trade secret. However, none of our contacts was aware of the analytic portfolio management technique like the one we propose in this paper. In this context, we have no baseline method used in practice to compare with. We therefore resort to simulating the application of the portfolio to the Google AdWords market by assuming that the estimates provided by Google AdWords for click number and CPC are reliable, *i.e.* we will assume that during the days after the decision time t_0 the number of views and the CPC for each adword will be as returned by Google AdWords and the evaluation of a portfolio will be derived using these data.

A critical aspect of our evaluation is the reliability on statistics as estimated (provided) by Google. As we are not acting directly on the Google AdWords market we have no way of verifying Google's data reliability. We thus take a conservative approach by setting the daily budget and max CPC bids to the maximum values allowed by Google in order to obtain higher bounds on CPC and click numbers. This ensures that the derived ROI represents a lower bound.

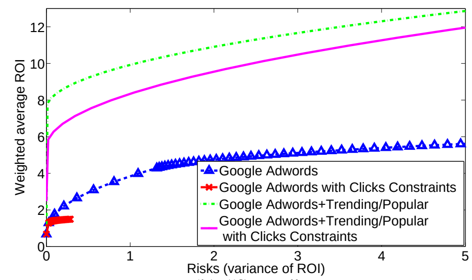
As we have discussed in Section II-C, the final CPC of an ad also depends on the Quality Score (QS), *i.e.* the relevance of the ad content to the adword. We use β, the click number constraint that is also defined in Section II-C, to consider the effect of QS. In detail, in order to account for the QS variations of trending and popular topics that are likely to be less relevant to the ad at the beginning of the campaign (time t_0) than the adwords suggested by Google, we conservatively set the number of achievable clicks on trending and popular topics to $\beta S^g_{t_0}(i,a)$, where $S^g_{t_0}(i,a)$ is the number of clicks of adword i in the ad a reported by Google AdWords at t_0. This assumption ensures that the obtained ROIs are likely to be lower than the actual values that would be observed in practice. In our current design, $\beta = 0.5$.

In our experiments, each ad at most has 10 adwords (as suggested by Google [1]), *i.e.* $|K(a)| \leq 10$ and the customer demand curve is compatible with a CED with $\alpha = 2.5$ (α is the price sensitivity coefficient). The PPC charged to the customer can be derived from Eq. 8 as $P(a) = 1.67\overline{\overline{CPC^+(a)}}$ to ensure maximal profit for broker.

In order to define and set a practical scenario we assume that the demand of the customer for his campaign is 500 clicks per day (this number of clicks per day is in accordance with values reported in [3]). In the forthcoming we apply the portfolio management approach developed in Section II-C to derive the efficient frontiers for the reference portfolio along with the augmented portfolio.



(a) "Candy"



(b) "Sports"

Fig. 2. The efficient frontier of two specific scenarios

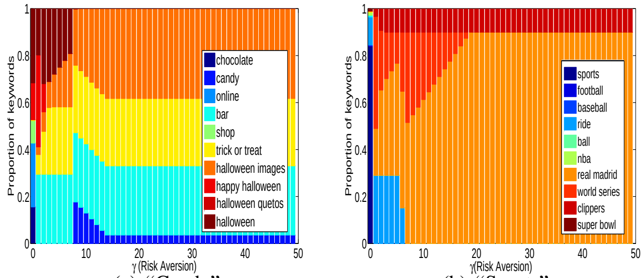
B. Portfolio performance analysis

We apply the above methodology to the two toy examples (Candy and Sport apparatus stores) and also to the random initial portfolios.

1) Two ad cases studies: We first derive $P(a)$ price for Candy ad (*resp.* Sports ad). The reference portfolio achieving the 500 clicks per day demand with the lowest risk (derived as section IV-A) attains $\overline{\overline{CPC^+(a)}} = 3.70$ USD per click for Candy ad (*resp.* 4.12 USD per click for Sports ad). This results in a selling price of $P(a) = 6.18$ USD per click for Candy ad (*resp.* 6.88 USD for Sports ad).

We show in Fig. 2 four efficient frontier curves for the two specific scenarios respectively, depicting the largest expected ROI for a given level of risk. We derive the efficient regions and the ROI respectively with and without the click number constraint, *i.e.*, without guarantee to attain the target click number resulting in the classical Markowitz portfolio case. The region of achievable (ROI, risks) pairs for which there exists a portfolio that can achieve this ROI with the given risk, is the set of points on the right and below the efficient frontier curve.

As expected the click number constraint reduces the achievable region area. For instance, in Candy scenario, the ROI of portfolio using Google AdWords with click number constraint can only reach 3.92 with a risk of 2.42, while the largest ROI of the portfolio using Google AdWords without click number constraints is 5.03 with a risk of 3.93. The constrained Google AdWords portfolio of Sports apparatus ad spans a very small



(a) "Candy" (b) "Sports"

Fig. 3. The portfolio composition of the two specific scenarios

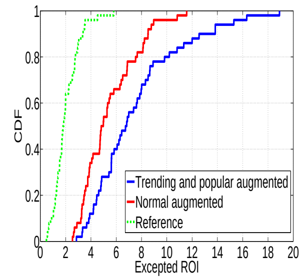


Fig. 4. CDF of $\overline{\overline{R(A)}}$, $\overline{\overline{R(B)}}$ and $\overline{\overline{R(C)}}$

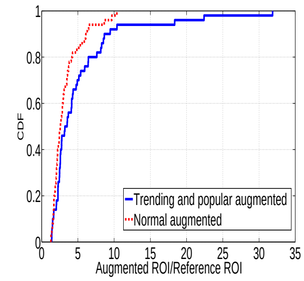


Fig. 5. CDF of $\frac{\overline{\overline{R(B)}}}{\overline{\overline{R(A)}}}$ and $\frac{\overline{\overline{R(C)}}}{\overline{\overline{R(A)}}}$

range of ROI as the largest ROI is 1.49 with a risk of 0.28. This can be explained by the additional restrictions the click number constraint brings to the optimisation model as there are only lesser number of adwords that can provide enough clicks to achieve the necessary demand.

Nonetheless the trending and popular topics largely extend the reachable region by augmenting the portfolio with adwords that seem to be more likely to meet the click number constraint. For example, in the Candy ad, to achieve the same ROI of 3.92, the augmented portfolio experiences a risk of only 0.34 while this risk of using Google AdWords is as high as 2.42. The augmented portfolio can even achieve a ROI as large as 5.95 but with an associated risk of 8.20.

We further show in Fig. 3 the keywords composition of the optimal portfolios for different values of risk aversion γ where the risk level is equal to 1. For lower values of γ, the portfolio contains a larger share of adwords suggested by Google to benefit from the average risk reduction effect of portfolios. With larger values of γ, the portfolio evolves towards a larger share of trending and popular topics (referring to Table III). The resulting portfolios from Fig. 3 also show that despite the possibility of utilising the 10 adwords in the augmented portfolio, all efficient portfolios just use 3 or 4 adwords. For example, we find that for a large range of risk aversion only 3 adwords ("real madrid", "clippers" and "world series") remain active in the "Sports apparatus" scenario. This shows that the higher performance of the augmented portfolio is not only due to the mechanical effect of the adword augmentation, but rather to the quality of the additional adwords.

Note that in practice we need to re-estimate the average and the variance of CPC and click number on a daily basis. There is then a need to "update" the parameters of the optimisation model. We note that the optimisation process for a small portfolio (*e.g.* our two examined scenarios) is executed in less than 2 seconds, so the time cost is not a major issue.

2) Random Portfolios analysis: Next we generalise our evaluation to random initial portfolio selection, a technique frequently used in stock market studies. Although in practice portfolios are not built randomly, a random portfolio can be considered to represent a particular case of portfolio built by the conscious action of a broker [10]. We first build a reference portfolio containing 5 adwords chosen randomly from all topics we have in the three Twitter datasets. In order to ensure that this portfolio is feasible (*i.e.* it can satisfy the customer's constraint), we check if the sum of the number of clicks in the portfolio can eventually reach the target demand per day. If the random portfolio is not feasible, we add one other randomly chosen adword till the resulting portfolio becomes feasible. This results in the "*reference portfolio*" called portfolio A. Next, we build two "*augmented portfolios*". The first augmented portfolio B is generated by adding randomly chosen adwords coming from trending and popular topics to the reference portfolio, while the second augmented portfolio C is generated by adding to the reference portfolio adwords coming only from the normal topics. Again, we limit the size of the augmented portfolio to 10 adwords. For each of these three portfolios, we derive the maximal ROI for a risk of 1 and compare the resulting ROIs. In order to decrease the impact of the randomness in the adwords choice we have generated independently 100 times the random reference portfolios along with the two attached augmented ones.

We show in Fig. 4 three cumulative distributions: the CDF of the $\overline{\overline{R(A)}}$, $\overline{\overline{R(B)}}$ and $\overline{\overline{R(C)}}$ obtained on each class of portfolio. We can observe that the two augmented CDFs are clearly on the right side of the reference one, showing that augmenting the portfolio by both ways can improve the ROI.

In order to determine which augmentation is more profitable, we calculate the ratios of the ROI achieved by the two augmenting strategies normalised by the ROI of the reference portfolio for a risk of 1. We show in Fig. 5, the CDF of the two ratios $\overline{\overline{R(B)}}/\overline{\overline{R(A)}}$ and $\overline{\overline{R(C)}}/\overline{\overline{R(A)}}$. We observe that both augmenting methods achieve a ratio that is always larger than 1, confirming that augmenting the portfolio always increases the achievable ROI. Moreover, the CDF curve for $\overline{\overline{R(B)}}/\overline{\overline{R(A)}}$ is on the right side of the CDF of $\overline{\overline{R(C)}}/\overline{\overline{R(A)}}$, meaning that the $\overline{\overline{R(B)}}$ is consistently larger than the $\overline{\overline{R(C)}}$. In particular, the average of $\overline{\overline{R(B)}}/\overline{\overline{R(A)}}$ is 5.20, meaning a $4.2\times$ improvement, while the average of $\overline{\overline{R(C)}}/\overline{\overline{R(A)}}$ is 3.27 and average of $\overline{\overline{R(B)}}/\overline{\overline{R(C)}}$ is 1.55.

V. Related Work

Several works have targeted advertisers and emphasised keyword optimisations. The authors in [14] propose an adword suggestion method exploiting semantic knowledge. Ghose *et al.* [16] study the relationship between adword characteristics, position of the advertisement and the search engine's ranking decision. A multiword adword recommendation system is developed in [27] reveals that specific text patterns can lead to high CTR in SEM. Other works have analysed the user behavior in SEM and proposed mechanisms to maximise the revenue [17][8][15]. However, most of these target the constitution of the portfolio and no one analyses the performance of the obtained portfolio and the way to optimise portfolio with Twitter, as we do.

As to the economy analysis of Internet, Zheng *et al.* [32] propose an optimisation model for the second market of mobile data, where data is traded between individual users. Chen *et al.* [13] model the financial aspects of 4G network deployment. Wu *et al.* [28] study the bundling sale strategy in online service markets. Hande *et al.* [18] investigate the pricing strategy of Internet connectivity services based on the alpha-fair utility model.

Online social networks have been explored for predication and optimisation in various scenarios. Zhang *et al.* [31] find that emotional tweet percentage significantly negatively correlated with Dow Jones, NASDAQ and S&P 500. The analysis in [9] indicates that the accuracy of DJIA predictions can be significantly improved by the inclusion of specific public mood dimensions in Twitter but not others. The authors in [25] find the sentiment of tweets to be associated with abnormal stock returns and message volume to predict next-day trading volume. The results in [11] confirm that trending topics offer a comparable visibility to the aforementioned traditional advertisement. Xu *et al.* [29] show the potential of online social networks in content objects popularity prediction.

VI. Conclusion and Future Works

Through an economic analysis of the third-party market we developed a portfolio management framework that controls the tradeoff between the Return On Investment and the risk resulting from uncertainty on current CPC and achievable click number in a search engine marketing context. To the best of our knowledge, this is the first work that models the economy of efficient portfolios of adwords. We studied the benefits of an efficient portfolio management, and in particular of the Efficient Frontier for comparing different portfolios. We also proposed to use trending and popular topics extracted from Twitter to augment the adwords portfolios. Our evaluation shows that the adwords augmentation is likely to improve the ROI on average by up to 4.2 times compared to a reference portfolio with the same level of risk. Even though in this paper we consider our model's application from a broker's perspective, the results obtained here are also relevant for an advertiser acting himself as the broker for his own ads. We believe this opens ways for further researches investigating rational management of adwords portfolio.

Acknowledgment

This work was supported by the National Basic Research Program of China No. 2012CB315801, the NSFC No. 61133015, 61272473 and 61572475.

References

[1] Adwords help for keyword setting. https://support.google.com/adwords/answer/2453981.
[2] Quality score. https://support.google.com/adwords/answer/2454010.
[3] Spyfu statistics. http://www.spyfu.com/, 2013.
[4] Google keywordplanner. https://adwords.google.com/ko/KeywordPlanner/, 2015.
[5] Google stop words list. https://sites.google.com/site/kevinbouge/stopwords-lists, 2015.
[6] C. Anderson. The long tail. *Wired magazine*, 12(10):170–177, 2004.
[7] M. S. Bernstein, J. Teevan, S. Dumais, D. Liebling, and E. Horvitz. Direct answers for search queries in the long tail. In *Proceedings of the SIGCHI*. ACM, 2012.
[8] A. Bhalgat, J. Feldman, and V. Mirrokni. Online allocation of display ads with smooth delivery. In *Proceedings of the 18th ACM SIGKDD*, pages 1213–1221. ACM, 2012.
[9] J. Bollen, H. Mao, and X. Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, 2011.
[10] P. Burns. Performance measurement via random portfolios. *Ssrn Electronic Journal*, 2004.
[11] J. M. Carrascosa, R. González, R. Cuevas, and A. Azcorra. Are trending topics useful for marketing?: visibility of trending topics vs traditional advertisement. In *Proceedings of the 1st COSN*. ACM, 2013.
[12] A. Chaabane, G. Acs, and M. A. Kaafar. You are what you like! Information leakage through users' Interests. In *Proceedings of the 19th Annual Network & Distributed System Security Symposium*, Feb. 2012.
[13] Y. Chen, L. Duan, and Q. Zhang. Financial analysis of 4g network deployment. In *INFOCOM, 2015 Proceedings IEEE*, 2015.
[14] Y. Chen, G.-R. Xue, and Y. Yu. Advertising keyword suggestion based on concept hierarchy. In *Proceedings of the 1st WSDM*. ACM, 2008.
[15] Y. Chen and T. W. Yan. Position-normalized click prediction in search advertising. In *Proceedings of the 18th ACM SIGKDD*, pages 795–803. ACM, 2012.
[16] A. Ghose and S. Yang. An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management Science*, 55(10):1605–1622, 2009.
[17] G. Goel and A. Mehta. Online budgeted matching in random input models with applications to adwords. In *Proceedings of the 9th SODA*. ACM-SIAM, 2008.
[18] P. Hande, M. Chiang, R. Calderbank, and J. Zhang. Pricing under constraints in access networks: Revenue maximization and congestion management. In *INFOCOM, 2010 Proceedings IEEE*. IEEE, 2010.
[19] P. W. H. C. IAB. Internet advertising revenue report 2014 full year results, 2015.
[20] A. Joshi and R. Motwani. Keyword generation for search engine advertising. In *Proceedings of the 6th ICDM Workshops*. IEEE, 2006.
[21] H. Markowitz. Portfolio selection. *The journal of finance*, 7(1):77–91, 1952.
[22] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking (ToN)*, 8(5):556–567, 2000.
[23] T. Preis, H. S. Moat, and H. E. Stanley. Quantifying trading behavior in financial markets using google trends. *Scientific reports*, 3, 2013.
[24] B. Skiera, J. Eckert, and O. Hinz. An analysis of the importance of the long tail in search engine marketing. *Electronic Commerce Research and Applications*, 9(6):488–494, 2010.
[25] T. O. Sprenger, A. Tumasjan, P. G. Sandner, and I. M. Welpe. Tweets and trades: The information content of stock microblogs. *European Financial Management*, 2013.
[26] S. Thomaidou and M. Vazirgiannis. Multiword keyword recommendation system for online advertising. In *Proceedings of ASONAM*. IEEE/ACM, 2011.
[27] T. Wang, J. Bian, S. Liu, Y. Zhang, and T.-Y. Liu. Psychological advertising: exploring user psychology for click prediction in sponsored search. In *Proceedings of the 19th ACM SIGKDD*, pages 563–571. ACM, 2013.
[28] W. Wu, R. Ma, and J. Lui. Exploring bundling sale strategy in online service markets with network effects. In *INFOCOM, 2014 Proceedings IEEE*, 2014.
[29] J. Xu, M. van der Schaar, J. Liu, and H. Li. Timely video popularity forecasting based on social networks. In *INFOCOM, 2015 Proceedings IEEE*, 2015.

[30] S. Yuan, J. Wang, and X. Zhao. Real-time bidding for online advertising: measurement and analysis. In *Proceedings of the Seventh International Workshop on Data Mining for Online Advertising*, page 3. ACM, 2013.
[31] X. Zhang, H. Fuehres, and P. A. Gloor. Predicting stock market indicators through twitter "i hope it is not as bad as i fear". *Procedia-Social and Behavioral Sciences*, 26:55–62, 2011.
[32] L. Zheng, C. Joe-Wong, C. W. Tan, S. Ha, and M. Chiang. Secondary markets for mobile data: Feasibility and benefits of traded data plans. In *INFOCOM, 2015 Proceedings IEEE*, 2015.