Online User Behavioural Modeling with Applications to Price Steering

Van Tien Hoang[1] and **Vittoria Cozza**[2] and **Marinella Petrocchi**[3] and **Rocco De Nicola**[1]

Abstract. Price steering is the practice of "changing the order of search results to highlight specific products" and products prices. In this paper, we show an initial investigation to quantify the price steering level in search results shown to different kind of users on Google Shopping. We mimic the category of *affluent* users. Affluent users visit websites offering expensive services, search for luxury goods and always click on the most costly items results at Google Shopping. The goal is checking if users trained in specific ways get different search results, based on the price of the products in the results. Evaluation is based on well known metrics to measure page results differences and similarities. Experiments are automised, rendering large-scale investigations feasible. Results of our experiments, based on a preliminary experimental setting, show that users trained on some particular topics are not always influenced by previous search and click activities. However, different trained users actually achieve different search results, thus paving the way for further investigation.

1 Introduction

Popular e-commerce websites, such as the Amazon Marketplace, offer a window to thousands of merchants, able to advertise their goods and services to millions of potential buyers.

Recently, the traditional advertising approach has moved towards a targeted one: the ad is shown only to online users with a specific profile - location, gender, age, e-shopping history are among the monitored aspects. This way, the merchant pays only for ads shown to users matching the ideal buyer for the merchant products. Targeted advertising is possible since the ads system is able to build a user profile tracking her online behaviour, e.g., on the e-commerce website, plus considering the data inserted by the user on the platform, at registration time.

Although personalised ads have the significant advantage to guide the customer mostly towards products she likes, concerns were born since the ads system could 1) hide to the user other potential interesting products [20]; and 2) expose user private information [3].

Price steering refers to the practice of changing the order of search results to highlight specific products prices [18]. In this work, we aim at studying if e-commerce websites rely on user past online behaviour to show her different product prices. In particular, we focus on Google Shopping, to discover if Google shows products of different price based on the *user willingness to pay*.

Google Shopping[4] is a promising platform to study the effect of price personalization. It allows vendors to reach a large number of customers, really interested in specific products, thus showing the right product to the right customer. Google Shopping creates a selling campaign, placing specific products "in front of millions of online shoppers searching on Google.com"[5]. This is possible since Google can access several information on the user search activity, not only including that on Google Shopping. Actually, Google monitors the circle of websites known as Google Display Network (GDN), a large set of websites publishing Google ads[6].

As shown in [13], Google builds *ad user profiles*, monitoring and learning behaviors when the users navigate on the GDN websites. Among the elements considered to build the ad profiles, there are the list of the visited websites, their topics, the time spent on each website, the number of times the user went to the website, the device the user is using for accessing at the platform, geo-localization of the user IP address.

Past work showed that price steering is affected by the user location, see [10, 19] and by the user device, as for the case of the online travel agency Orbitz [17]. Orbitz realised that Mac users were more interested in costly hotels and traveling services than Windows users. Consequently, the agency showed the most costly results as the first results for Mac users.

Our work focuses on how the user behaviour, e.g., visiting a website of luxury goods, clicking on expensive products, affects the price of the items shown as the result of future queries over Google Shopping. We emulate the on-line behaviour of an *affluent* user and we compare her results list with one of a fresh user, which has not searched before on Google Shopping. Preliminary results show that, overall, affluent profiles have been shown different results with respect to those shown to the fresh control user. However, there is no a fixed rule, leading first to the most expensive products shown to the affluent users. The difference in the results list is however worth to be acknowledged, and calls for further investigation, with a more complex experimental settings and a more extensive evaluation.

The rest of the paper is as follows. Next section briefly presents related work in the area. Section 3 describes our methodology. In Section 4, we describe the experiments and we give the results. Finally, Section 5 concludes the paper.

2 Related Work

This section briefly relates on literature in the area of personalization of web results, particularly focusing on price steering and price

[1] IMT School for Advanced Studies Lucca, Italy, email: {vantien.hoang, rocco.denicola}@imtlucca.it

[2] Electrical & Information Engineering Department (DEI), Polytechnic University of Bari, Italy, email: vittoria.cozza@poliba.it

[3] IIT-CNR, Pisa, Italy, email: m.petrocchi@iit.cnr.it

[4] http://www.google.com/shopping

[5] https://services.google.com/fh/files/misc/product_listing_ads_intro.pdf

[6] https://support.google.com/adwords/answer/2404190?hl=en

discrimination. While price steering denotes the practice of showing different products with different prices to different users, discrimination is a similar practice, but related to the same product. Work in [18] gives an alike definition of price steering: a scenario in which e-commerce websites show to the unaware user different search results, based on the user willingness to pay (defined as the maximum amount of money a customer is willing to spend for a product).

Mikians et. al [19] detected online price discrimination by collecting data from 340 real Internet users over 18 countries. The analysis focused on how the price of the same product, offered from a set of retailers, varies retailer per retailer. Outcome denotes geographic location as the main factor affecting the prices.Work in [10] analysed price discrimination by adopting fictitious users, mimicking a visit to shopping websites, from 6 different locations, for 7 days. Even in this case, results show that user location has an impact on price discrimination.

In [12], the authors extensively measures both price steering and discrimination. With both real data collected through Amazon Mechanical Turks and synthetic data from controlled experiments using a headless web browser[7], the authors analyse the prices offered by a plethora of online vendors. The work finds evidence of price differences by different merchants and retailers: their websites record the history of clicked products to discriminate prices among customers.

The issue of price steering analysed in this paper has close relationships with the ties between online user behaviours and the search results (and/or advertisements) presented by a search engine (or a website) to the same user. Indeed, price steering and discrimination constitute only one aspect of a wide phenomenon, originally put in the spotlight by Pariser with his Filter Bubbles [20] and investigated by seminal work on web search personalization, like, e.g., the one in [11]. Particularly, online user behaviour has been investigated widely in relation to targeted advertising. As an example, work in [16] reveals how Google ads are selected for specific users, according to their activities over the Internet. Experimental results in the paper claim that 65% of ads categories shown to users have been targeted according to their behaviours. Targeted ads have been studied also in [2], achieving results consistent with [16]. Recently, this targeting phenomenon has been investigated in [15] on ads in Gmail, resulting in an evidence of linking users behaviours and shown ads also on the email service. Overall, tracing online behaviour is commonly adopted - and such information is commonly exchanged among websites - to determine which ads are shown to users.

In this work, we take inspiration from the analysis of online user behaviour to evaluate the effect on that behaviour on product prices shown to the user.

3 Methodology

Our goal is measuring how the user online behaviour affects price steering on Google Shopping. We use the approach which is similar in [9, 12, 21]. At first, we consider users of two categories: *affluent* and synthesized *control* users. Intuitively, the former feature higher willingness to pay than the latter. Thus, aiming at mimicking their behaviour, we assume that affluent users search and click on more expensive products than control users. We have considered two kinds of behaviours: 1) visiting web pages, and 2) searching for keywords on Google Shopping. Visiting a page means staying on that page for a while and also scrolling the page. We define the affluent user behaviour as follows:

[7] http://phantomjs.org

- an affluent user visits websites selling luxury goods;
- an affluent user searches for keywords representing luxury goods on Google Shopping and
  - she visits the three results representing the first three products whose price is above the average (among all the obtained results)
  - she visits those result pages that are the same of a previously visited website selling luxury goods.

The behaviour of a control user is different, she is idling while the affluent is in action.

Our expectation is that, when users query Google Shopping after a training phase where they behave as described above, affluent (resp., control) ones will likely see the highest (lowest) price products ranked first in their list of results.

In order to identify websites and keywords for luxury goods, we have exploited a tool originally intended for setting up targeting advertisements, the Display Planner tool of Google AdWords. The tool guides the user to find websites and keywords inherent to specific topics and terms[8].

It is well known that Google monitors the users' behavioural activities over the Internet through tools such as Google Analytics, Google Plus, and the Google ads system[9]. Thus, we train two affluent user profiles according to the specific online behaviours described above. Each profile has a control user profile associated. A control user has the same configuration as the user she is associated to (same browser, same OS). The difference is that control users are not logged into a Google account. Then, we compare the results obtained searching on Google Shopping the same keywords for both the trained users and the control users. In detail, the training and test phase are as follows:

- Training step 1: The two affluent users visit a list of websites with topics related to their category. Websites have been chosen using the Google Display Planner.
- Training step 2: The two affluent users search on Google Shopping for keywords related to their user category (keywords have been chosen according to the Display Planner, too). Then, they click on the most expensive product results and on those results coming from websites visited at step 1 (when present).
- Test: All users (trained plus control) search for new products on Google Shopping. They do not interact with the results.

We let the two affluent users repeat the training phase eight times. At the end of each training, we run the test. This is for building a longer behavioral history of the user. Indeed, Google itself states that it uses browsing and search histories to personalise the results and enhance the user experience[10]. This is why we have repeated the training phase eight times, always on the same profile, to emphasize possible personalization aspects. One evidence of the efficacy of this modality is in [1]: Google infers the interests of users only after a certain amount of websites visiting. Another evidence is in [6], where the authors described that Google took five days of training on a single user to produce personalized news content.

[8] https://adwords.google.com/da/DisplayPlanner/Home
[9] https://www.google.com/policies/privacy/
[10] https://www.google.com/policies/terms/

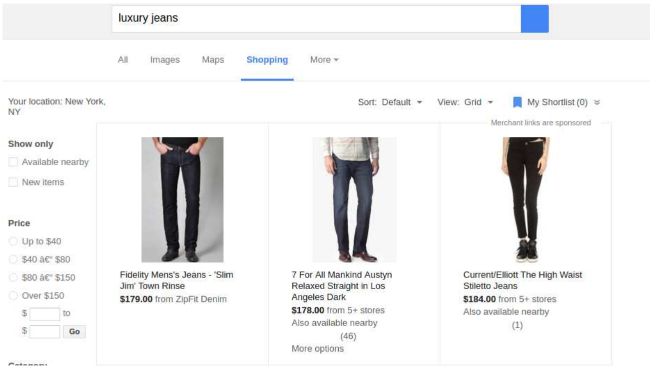

Figure 1. Top 3 results on Google Shopping: control, 3rd test block

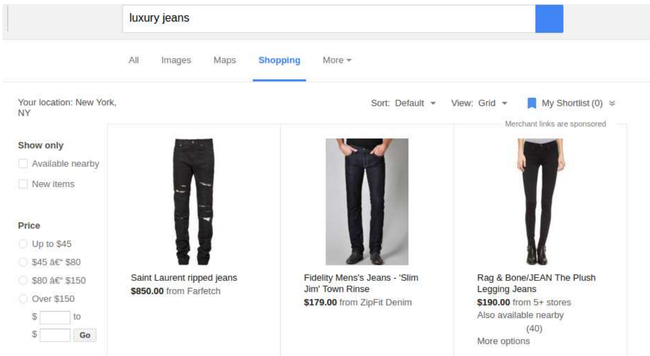

Figure 2. Top 3 results on Google Shopping: affluent, 3rd test block

4 Experiments

For our experiments, we use AdFisher [7], a freely available automated tool[11]. Natively, the tool functionalities allow to analyse interactions between online user behaviors, advertisements shown to the user, and advertisements settings. Later, AdFisher has been extended for handling Google searches and news searches, to measure personalization in query results, see, e.g., [6] and applied to novel news search experiments [5]. AdFisher has also been used for statistical evaluations, e.g., to measure how users are exposed to Wikipedia results, in return to their web searches, see, e.g., [4]. The interested reader can refer to [8] for a full survey on tools for measuring and analysing users' interactions with online services (including AdFisher).

In our work, AdFisher runs browser-based experiments that emulate search queries and basic interactions with the search results, e.g., interacting with those search results whose price is above or below the price average on the total of the results, or those results belonging to a list of previously visited websites. Github hosts our extended version[12].

AdFisher interacts with Selenium, a web browser automation tool. Selenium allows to run a unique instance of Firefox creating a fresh profile, with new associated cookies, the so called *Firefox profile*. The Firefox profile that is used is stripped down from what is installed on the machine, to only include the Selenium WebDriver.xpi plugin. Further, we take advantage of a plugin to automatically obtain the Python code for recording actions on web pages (e.g., clicking, typing, etc.), provided by Selenium IDE for Firefox[13]. All the experiments are done on the Firefox web browser version 43.0.4, controlled by Selenium in Python, under XUbuntu 14.04.

To simulate different real users, we browse from different IP addresses, implementing a solution based on SSH tunneling to remote computers. To simulate many computers from one geographic area, we use the VPS services of Digital Ocean[14]. It is well known that, when a user enters a query to Google, the query is unpredictably sent to many distributed servers, to retrieve the results. This could produce noise due to inconsistent data among different servers. To avoid the issue, we query only towards specific Google servers IP addresses, as in [11].

Finally, we use the following metrics to evaluate the search results of the test:

- Jaccard Index: given two sets P and Q, Jaccard Index is 1 when the sets are identical and 0 when their intersection is empty.
- NDCG (Normalized Discounted Cumulative Gain), measuring the similarity between a given list of results and the ideal list of results. Originally introduced as the non-normalised version DCG in [14], NDCG has been adopted in [12] for measuring price steering. For each result r, there is one gain score g(r), representing its price. For a result page $R = [r_1, r_2, ..r_k]$, we have $DCG(R) = g(r_1) + \sum_{i=2}^{k}(g(r_i)/log_2(i))$. NDCG is $DCG(R)/DCG(R')$, where R' is the ideal result page (a list in which the results are shown from the most expensive to the least one). We create R' by unionising the results returned, for the same query, to affluent and control profiles. Then, we sort such results from the most expensive to the least expensive one.

11 https://github.com/tadatitam/info-flow-experiments
12 https://github.com/tienhv/Adfisher_for_GoogleShopping
13 http://www.seleniumhq.org/projects/ide/
14 https://www.digitalocean.com/

4.1 Settings and results

We have extended the AdFisher functionalities to handle Google Shopping pages and to mimic the behaviours described in Section 3. We automatically implement the whole experiments for the affluent and control users. The extended AdFisher also stores the query results for further analysis, as the calculation of the NCDG metric.

For the training phase, we emulate two user profiles logged into Google. We consider 80 websites for each type of user profile and we let the profiles visit all of them. Such choice has been driven by [1], which proved that visiting 50 websites is enough for Google Ads to infer the user interests. The website visit time for a user is a random value, however less than 30 seconds (such threshold being estimated following the `alexa.com` statistics). Furthermore, each profile has been trained with 15 training keywords, and 3 were the resulting links to be clicked, associated to the top 3 most expensive products. Table 1 shows an excerpt of the visited websites and the searched keywords, selected with the help of the Google Display Planner.

For the test phase, we still consider the two trained profiles, plus two control ones. All feature the same behaviour, which consists of querying Google Shopping with the same test keywords. Then, we collect the results. Figure 1 and Figure 2 show the first results showed to affluent and control users, for a specific test query. We extract links and prices of all the results to calculate the metrics listed in Section 3.

The training and the test phases are repeated eight times per user. Each session lasts around 90 minutes.

Table 2 shows NCDG values for each test session (results are for query "luxury shoes"), for each user.

Figures 3 and 4 plot, resp., the Jaccard index and the Kendall index, for the eight test sessions about "luxury shoes". The blue lines represent are calculated over the results pages of *Affluent 1* and *Control 2* users, while red lines are calculated over the result pages of *Affluent 2* and *Control 1* users. This is to consider two users trained in same ways and connected from two different machines. Jaccard index shows evidence of results customization, while Kendall index says that, most of the times, the results for affluent and control profiles have a level of agreement (featuring a positive values for that index).

Figure 5 plots the values obtained calculating the average NCDG of the two affluent users and the two control users, over the eight test sessions. The test query is "luxury shoes". Average NCDG indicates that even if affluent and control users follow the same pattern of pricing ordering, the former is closer to the ideal list results (where the most expensive products are in the first positions).

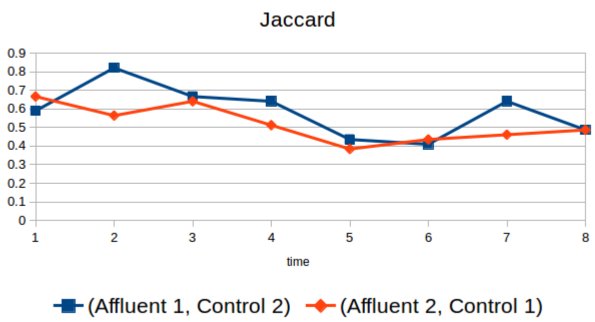

Figure 3. Jaccard Index for "luxury shoes"

Figure 6 shows the average NCDG value for all the test queries (averaged both per profile and over the eight sessions). The Figure

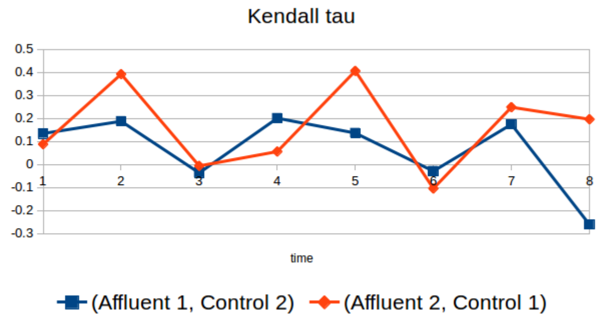

Figure 4. Kendall Index for "luxury shoes"

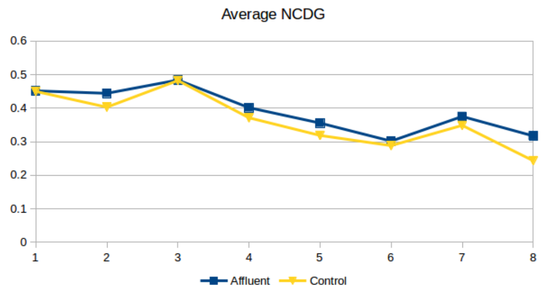

Figure 5. Average NCDG for "luxury shoes"

shows, at a glance, how, in some cases, NCDG values are higher for affluent users than for control. The control users almost always obtain NCDG values lower than the affluent (or comparable in those cases with very similar values for the two kind of profiles). However, we argue that results are also affected by the specific search query over Google Shopping. As an example, Figure 7 shows the NCDG values for the test query "luxury jeans" for affluent 1 and the average of results of the two controls, over the eight sessions. The picture clearly indicates NCDG values that are greater for the affluent user. Instead, "mens dress casual shoes" (not shown in a Figure over the eight sessions) provides a higher value for the control which is an opposite result with respect to our expectations. While these initial results are promising, there is the need of further evaluation, where more, and more generic keywords, should be tested, at a larger scale.

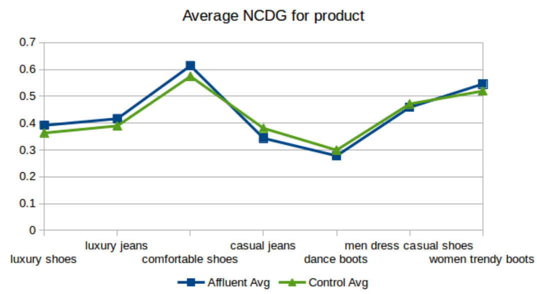

Figure 6. NCDG values averaged over profiles and sessions, per product

It is worth noting that, as introduced and motivated at the end of Section 3, the two affluent users (as well as the two control ones) are identical in terms of behaviour. Further, browser and OS settings are the same, while the IP address from which they browse is different

Table 1. Training phase: An excerpt of websites and keywords

training websites	training keywords	test keywords
outfitideashq.com seriousrunning.com lululemonmen.com storelocate.us, skateboardingmagazine.com haircutinspiration.com	mens fashion shoes, trendy boots, formal shoes for men, jogging shoes, cheap designer shoes, athletic shoes	mens dress casual shoes, luxury shoes, dance boots, comfortable shoes, women trendy boots, luxury jeans, casual jeans

Table 2. NCDG of "luxury shoes" for 8 test sessions

luxury shoes	slot 1	slot 2	slot 3	slot 4	slot 5	slot 6	slot 7	slot 8
Affluent 1	0.45	0.45	0.49	0.40	0.38	0.32	0.37	0.33
Control 1	0.45	0.42	0.48	0.38	0.32	0.30	0.35	0.24
Affluent 2	0.45	0.44	0.47	0.40	0.33	0.28	0.38	0.30
Control 2	0.45	0.39	0.48	0.36	0.32	0.28	0.34	0.25

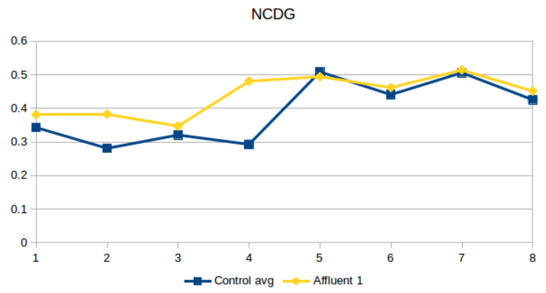

Figure 7. NCDG per session, for "luxury jeans"

(different devices, from the same geographical area). We have not compared directly the two affluents, since their results pages could be different for uncontrollable effects, such as timeouts or network delays (we are indeed emulating different IP addresses). The existence of sources of noise is also the reason why we have chosen to show, for some experiments, the average of the results of the two users.

5 Conclusions

We have designed and implemented a methodology to train and test user behaviours on Google Shopping, for evaluating a potential price steering, based on the *willingness to pay* attitude of the users. We have analysed the results list of affluent and control users. Affluent users were trained over eight training sessions. The results lists were obtained over eight test sessions, one at the end of each training session. The outcome of the experimentation is that, for most of the test queries, the result list of the affluent user is biased towards more expensive products than the one of the control user. However, the experiments results pave the way for further investigation. Indeed, we can imagine to 1) mimic queries from different geographical areas (not considered here, but recognised by past work as an impact factor for price manipulation); 2) use location via IP address as the major measurement, instead of artificial user profiles, because in some countries (like USA), the location/postcode is a strong indicator for economic situation, religion and race; 3) augment the number of training and test queries; 4) expand the duration of each training and test experiment; 5) mimic queries by different kind of users, e.g., mimic *budget* users, which always search for cheap products and services. Our experimental approach is general enough to be applicable to other e-commerce websites, like, e.g., *Amazon.com* and *eBay.com*.

6 Acknowledgment

This research has been partially funded by the Registro.it project *MIB* (My Information Bubble).

REFERENCES

[1] Barford, P., Canadi, I., Krushevskaja, D., Ma, Q., Muthukrishnan, S.: Adscape: Harvesting and analyzing online display ads. In: Proceedings of the 23rd International Conference on World Wide Web. pp. 597–608. WWW '14, ACM (2014)

[2] Carrascosa, J.M., et al.: I Always Feel Like Somebody's Watching Me Measuring Online Behavioural Advertising. In: 11th Emerging Networking EXperiments and Technologies. ACM (2015)

[3] Conti, M., Cozza, V., Petrocchi, M., Spognardi, A.: TRAP: using targeted ads to unveil google personal profiles. In: IEEE International Workshop on Information Forensics and Security. pp. 1–6 (2015)

[4] Cozza, V., Hoang, V.T., Petrocchi, M.: Google web searches and Wikipedia results: a measurement study. In: Proc. of 7th Italian Workshop on Information Retrieval (IIR) 2016. CEUR Workshop Proceedings (2016)

[5] Cozza, V., Hoang, V.T., Petrocchi, M., Spognardi, A.: Experimental measures of news personalization in Google News. In: SoWeMine 2016: 2nd International Workshop on Mining the Social Web. Springer (2016)

[6] Datta, A., Datta, A., Jana, S., Tschantz, M.C.: Poster: Information flow experiments to study news personalization. In: Computer Security Foundations Symposium (CSF), 2015 IEEE 28th. IEEE (2015)

[7] Datta, A., Tschantz, M.C., Datta, A.: Automated experiments on ad privacy settings. Proceedings on Privacy Enhancing Technologies 2015(1), 92–112 (2015)

[8] Englehardt, S., Narayanan, A.: Online tracking: A 1-million-site measurement and analysis (May 2016), [Technical Report]

[9] Englehardt, S., et al.: Web privacy measurement: Scientific principles, engineering platform, and new results. Manuscript posted at http://randomwalker.info/publications/WebPrivacyMeasurement.pdf (2014)

[10] Guha, S., Cheng, B., Francis, P.: Challenges in measuring online advertising systems. Internet Measurement Conference pp. 81–87 (2010)

[11] Hannak, A., al.: Measuring personalization of web search. In: 22nd World Wide Web. pp. 527–538 (2013)

[12] Hannak, A., al.: Measuring price discrimination and steering on e-commerce web sites. In: IMC. pp. 305–318. ACM (2014)

[13] Haveliwala, T., Jeh, G., Kamvar, S.: Targeted advertisements based on user profiles and page profile (Nov 27 2012), US Patent 8,321,278

[14] Järvelin, K., Kekäläinen, J.: Cumulated gain-based evaluation of ir techniques. ACM Trans. Inf. Syst. 20(4), 422–446 (Oct 2002)

[15] Lecuyer, M., Spahn, R., Spiliopolous, Y., Chaintreau, A., Geambasu, R., Hsu, D.: Sunlight: Fine-grained targeting detection at scale with statistical confidence. In: Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security. pp. 554–566. CCS '15, ACM (2015)

[16] Liu, B., Sheth, A., Weinsberg, U., Chandrashekar, J., Govindan, R.: AdReveal. In: HotNets-XII. pp. 1–7. ACM (2013)
[17] Mattioli, D.: On Orbitz, Mac users steered to pricier hotels. Wall Street Journal (2012), http://on.wsj.com/LwTnPH
[18] Mikians, J., Gyarmati, L., Erramilli, V., Laoutaris, N.: Detecting price and search discrimination on the Internet. In: Proceedings of the 11th ACM Workshop on Hot Topics in Networks. pp. 79–84. HotNets-XI, ACM (2012)
[19] Mikians, J., et al.: Crowd-assisted search for price discrimination in e-commerce: First results. CoNEXT pp. 1–6 (2013)
[20] Pariser, E.: The Filter Bubble: What the Internet is hiding from you. Penguin UK (2011)
[21] Tschantz, M.C., Datta, A., Datta, A., Wing, J.M.: A methodology for information flow experiments. CoRR abs/1405.2376 (2014)