Transparency in Keyword Faceted Search: An Investigation on Google Shopping

Vittoria Cozza[1(✉)], Van Tien Hoang[2], Marinella Petrocchi[3], and Rocco De Nicola[2]

[1] Department of Information Engineering, University of Padua, Padua, Italy
covitti@dei.unipd.it
[2] IMT School for Advanced Studies, Lucca, Italy
{vantien.hoang,rocco.denicola}@imtlucca.it
[3] IIT Institute of Informatics and Telematics, National Research Council (CNR), Pisa, Italy
marinella.petrocchi@iit.cnr.it

Abstract. The most popular e-commerce search engines allow the user to run a keyword search, to find relevant results and to narrow down the results by mean of filters. The engines can also keep track of data and activities of the users, to provide personalized content, thus filtering automatically out a part of the results. Issues occur when personalization is not transparent and interferes with the user choices. Indeed, it has been noticed that, in some cases, a different ordering of search results is shown to different users. This becomes particularly critical when search results are associated with prices. Changing the order of search results according to prices is known as price steering. This study investigates if and how price steering exists, considering queries on Google Shopping by users searching from different geographic locations, distinguishable by their values of Gross Domestic Product.

The results confirm that products belonging to specific categories (e.g., electronic devices and apparel) are shown to users according to different prices orderings, and the prices in the results list differ, on average, in a way that depends on users' location. All results are validated through statistical tests.

Keywords: Keyword faceted search · Information retrieval · Personalisation · Price steering · Automatic browser interactions · Permutation tests

1 Introduction

Popular e-commerce websites, such as *Amazon Marketplace*, *eBay* and *Google Shopping*, offer a window to thousands of merchants, providing products and services to millions of potential buyers [25,26].

Some of the most popular e-commerce websites let users search for products by simply issuing a keyword search. Then, a number of filters can be activated

P. Manghi et al. (Eds.): IRCDL 2019, CCIS 988, pp. 29–43, 2019.
https://doi.org/10.1007/978-3-030-11226-4_3

to constraint the search results. Such filters are usually defined over a number of attributes of the products, like the vendor, the brand, the size, and the price range. The filters allow to narrow the results list, and, together with characteristics of the users, such as their behavior, the location they search from, and the data on their profiles, allow to obtain a personalised set of items as the outcome of the search.

The dark side of personalization relies however in the fact that filters can be activated, or changed, without the user's awareness and consent. In this case, the search engine acts in a *not transparent* way. A recent work in [24] defines different kinds of discrimination possibly enacted by search engines, among the others the *user bias*, taking place when the values of some of the attributes that characterise the users, e.g., their race or gender, influence the results presented to them. Consequences of lack of transparency and bias are, e.g., to hide potentially interesting products [22], give relevance to some news with respect to others [7], expose different prices for the same product, depending, e.g., on the characteristics of the user making the search [18], and even reveal users' private information [4]. Recently, independent developers have started to design tools to avoid part of such consequences. As examples, there exist tools that remove users' personal information when sending an online request[1], escape from echo chambers[2], increase transparency of personalization algorithms[3].

This work considers the possible hidden actuation of a price filter by the search engines, based on the geographical location of the user that makes the research. One of the possible way in which such a filter can be actuated is to change the order of the results shown to the users according to their prices. This practice is known in the literature as *price steering* [12].

In particular, this study investigates if, and to what extent, the practice of price steering is actuated with respect to users in countries characterised by different Gross Domestic Product (GDP) values.

To minimise possible noise caused by different factors, the experiments do not consider user search behavior but only search location, based on the IP address of the users. The experiments are launched on Google Shopping US. Indeed, Google is well known for tracking users and providing them with personalised services (e.g., target advertising)[4]. Google Shopping let users perform Keyword and Facet Search to explore a large structured data collection: the retail products.

Google Shopping both associates attributes to the products and considers the information of the users' profiles to provide personalized results. Aiming at analysing if an untransparent match between products attributes and users' profiles is actually exploited, we collect the list of search results from different users in three distinct cities of three different countries. Given past literature results, highlighting the user' habits of mainly focusing on results that appear first (see, e.g., [11]), the analysis is limited to the products shown in the first

[1] http://www.debiasyourself.org/.
[2] https://www.escapeyourbubble.com/.
[3] https://facebook.tracking.exposed/.
[4] https://www.google.com/retail/shopping-campaigns/.

page of Google Shopping, corresponding to 40 items. Statistics on our results are computed by considering the top 3 and the top 10 items. Two kinds of outcome are taken into account: (i) the order in which the results are shown, in terms of prices, with respect to an ideal list where the prices are shown from the most to the least expensive ones, and (ii) the average prices shown in the results list. Both metrics lead to significant results. The main highlights of this work are as follows:

- when users search from US, Google Shopping US tends to show products ordered from the most to the least expensive, differently from searches from India and Philippines;
- considering the product category "electronic devices", and relying on an ideal list of results where the results are shown from the most to the least expensive one, the order of the electronic products shown to users searching from US is the most similar to that of the ideal list;
- the average price of the top ten electronic devices shown to users searching from US is lower than that for Philippines and India;
- considering the product category "body care", and still relying on the ideal list of results, the order of products shown to users searching from Philippines is the most similar to that of the ideal list;
- for the "body care" products category, the average price for products shown to users searching from Philippines is higher than that shown to users searching from US;

The results of the experiments are validated by means of statistical tests. To pave the way for further evaluations, even different from the ones presented in this work, the data collected during the experiments (i.e., the search results, for each product and for all the tested synthetic users) are publicly available[5].

The rest of the paper is organised as follows. Next section presents related work in the area. Section 3 describes the methodological approach, while Sect. 4 introduces the experiments and discusses their outcomes. Finally, Sect. 5 concludes the paper.

2 Related Work

The literature reports about two well-known practice of prices personalisation over the internet, i.e., "price steering" and "price discrimination", see, e.g., [12, 19]. While price steering denotes the act of changing the order of the results shown to the users according to their prices, price discrimination is the practice of offering different prices, to different users, for the very same product.

The concrete risk of price discrimination and steering was already considered more than one decade ago [21]. It was described how by relying on a large scale collection of personal information, like user behaviour and demographics, prices manipulation can be easily implemented in different scenarios. As a popular

[5] http://doi.org/10.5281/zenodo.1491557.

example, in August, 2012, the Wall Street Journal announced a real case of price steering [27]: an online travel agency, called Orbitz, was found to be steering Mac users towards more expensive hotels.

In [12], price steering and discrimination were extensively measured. With both real data collected through Amazon Mechanical Turks and synthetic data from controlled experiments using the non-GUI web browser in [13], the authors analysed the prices offered by a plethora of online vendors. The work found evidence of price differences by different merchants: their websites used to record the history of clicked products, to discriminate prices among customers.

In [18], the authors consider both price discrimination and price steering (the latter being referred as 'search discrimination). Collecting data from more than 200 online vendors, they did not find traces of price and search discrimination depending on the OSs, browsers, and their combination. Regarding price discrimination only, they discovered noticeable differences of products prices, particularly for the digital market (e.g., e-books, videogames) and depending from the kind of vendor (single e-shop *vs* aggregator of e-commerce websites). Finally, they ran experiments by synthetically varying the search history of users, thus building two kind of profiles (conscious user *vs* affluent user): the results showed a not negligible level of search discrimination.

In [19], the authors analysed real data collected from 340 internet users from 18 countries. The analysis focused on how the price of the same product, offered by a set of retailers, varied from retailer to retailer. The geographical location of users turned out to be one of the main factors affecting the prices differences, even if its constant influence over all the experiments was not assessed.

Work in [14] considered price discrimination actuated by popular accommodation websites. The authors experimented with different features: the user location and the system configuration. The study revealed that a price discrimination was indeed applied by booking providers according to locations, language and user agent (that contains information about the operating system and the browser).

According to [31], not all retailers adopt personalisation of prices. The authors monitored 25 airline websites for 3 weeks, with dozens of unique user profiles to analyze airline tickets. The experiments were automated by a headless Webkit browser (similar to that of [12,16]). The results did not reveal the use of any systematic price discrimination. Work in [6] collected price variations of search results over Google Shopping, varying the on-line behaviour of synthetic users: searched keywords, visited pages, clicked products on the visited pages; such research did not provide relevant evidence of price personalisation.

Work in [2] throws evidence on how sellers may set prices using so called dynamic pricing algorithms. While making vendors more competitive, such kind of algorithms may cause intentional agreements on prices among participants on the same side in a market, a practice known as price fixing [29], and cases exist of pricing algorithms pushing prices to unrealistic heights, like the unbelievable amount of $23.7 million for a scientific book about flies [3].

Regarding the synthesis of user profiles, work in [1,18,31] provides a useful way of constructing them based on OSs, web browsers, user behaviors and geographical locations. These investigations give useful references for further studies, including the present one.

As testified by the results in the literature, price manipulation on online markets is an actual issue. In particular, geographical areas have been already identified as an impact factor for price steering and discrimination. With respect to related work in the area, this research concentrates on price steering and it aims at identifying if, and how, such a practice is related to geo-economics features of the geographical areas from which users search on the internet. Previous work analysed the connection between online shopping habits and location and/or income of users (without however considering price steering practices). As an example, in [32], it has been shown a comparison among Chinese and Dutch people, assessing that the latter are more sensitive to advertisements proposing branded and more expensive products with respect to China.

Finally, in place of searching on specific retailers' websites, as in [12], this work focuses on Google Shopping, mainly because Google is well known to provide personalised services to the users. However, the experimental approach is general enough to be applicable to different search engine platforms.

3 Methodology

This work quantifies the differences in product prices, on search results in return to users queries on Google Shopping, US version. In particular, the considered scenario is one where users search from cities located in different countries, characterised by different values of Gross Domestic Product. All the experiments are conducted by collecting results during July, 2016. The prices are the retail ones displayed by Google Shopping in US dollars (thus, excluding shipping fees).

User Profiles. In personalization measurement studies, a very relevant aspect, often underestimated, is the quality of the user profiles [23]. This study simulates real user profiles surfing Google Shopping US from different geographic locations. The creation of a new profile corresponds to launch a new isolated web browser client instance and open the Google Shopping US web page. Also, instead of providing artificial ad hoc user profiles, the experiments consider the IP addresses of the users. Indeed, it has been shown that, in some countries (e.g., in US), the manual insertion of the user location (in the form, e.g., of a postal code) in the user profile do unfairly affects the price results [30]. Furthermore, locations and postal codes can be easily altered during the profile registration phase [17].

Emulating Users. To mimic real users, the synthetic users can browse, scroll pages, stay on a page, and click on links. A fully-fledged web browser is used to get the correct desktop version of the website under investigation. This is because websites could be designed to behave according to user agents, as witnessed by the differences between the mobile and desktop versions of the same website.

Several frameworks have been proposed for interacting with web browsers and analysing results from search engines. The interested reader can refer to [9] for a complete survey. This work considers tools able to run browser-based experiments to emulate search queries and basic interactions with the search result, and that could be easily extended with new user behaviours and new evaluation metrics. Both AdFisher[6] [8] and OpenWPM [9] meet these constraints. Past work on similar topics (see, e.g., [5–7]) experimented issues with AdFisher, mainly related to the recovery of browsers data after crashing. Thus, this research adopts OpenWPM, since it features a recovery mechanism after crashes (the experiments run, on average, 24 h). OpenWPM is automatised with Selenium[7] to efficiently create and manage different users with isolated Firefox and Chrome client instances, each of them with their own associated cookies.

In all the experiments, the software runs on our local server, but the browser's traffic is redirected to the designated remote servers (i.e., to India), via tunneling in SOCKS proxies. By operating in this way, a guarantee is that all the commands are simultaneously distributed over all the proxies. The experiments adopt the Mozilla Firefox browser (version 45.0) for the web browsing tasks and run under Ubuntu 14.04. Also, for each query, we consider the first page of results, counting 40 products. Among them, the focus of the experiments is mostly on the top 10 and top 3 results. We limit the investigation to the first results, following past studies that highlight the user's habits to concentrate only on them [11].

Metrics. To evaluate the result pages and quantify the differences in prices of the search results, this paper relies on a metric widely adopted in the information retrieval research area, namely the Normalized Discounted Cumulative Gain - NDCG - metric that measures the similarity between a given list of results and an ideal list of results. In this work, the ideal list of results is a list in which the products are listed from the most expensive to the least expensive one. This specific order is motivated by the fact that we are investigating if, and to which extent, the most expensive products are shown first to the user. In particular, the 'best' way to implement price steering would be to show products from the most expensive to the least expensive. Thus, this work will compare the results of the experiments with that ideal list.

NDCG, originally introduced in [15] in its non-normalised version DCG, has been already adopted in [12] for measuring price steering. For each search result r, there is one gain score g(r), representing its price. In a page with k results, we let $R = [(r_1), (r_2), ..(r_k)]$ and $R' = [(r'_1), (r'_2), ..(r'_k)]$, where r'_1 is the most expensive result and r'_k is the least expensive one. Thus, R' is the defined ideal list of results, obtaining:

$$DCG(R) = g(r_1) + \sum_{i=2}^{k}(g(r_i)/log_2(i))$$

$$NDCG = DCG(R)/DCG(R').$$

[6] https://github.com/tadatitam/info-flow-experiments.

[7] http://www.seleniumhq.org/.

Similar to [12], this research creates R' by first unioning the results returned for the same query to all the profiles under investigation, and then sorting such results from the most expensive to the least expensive one. For each query, this work calculates the NDCG obtained from the corresponding result page. After a profile successfully executes x queries, one NDCG vector is obtained, with x elements for it, where each element corresponds to the NDCG value of a single query. This is the *NDCG vector for a single profile*.

The *NDCG vector of a location* is instead obtained by unioning all the NDCG vectors of the profiles querying from that location. We consider more than one profile from the same location, to assure a higher confidence when evaluating the results.

As an example, with five profiles querying from the same location, there are five NDCG vectors associated to those profiles. The five vectors are then unioned into a single one (with length equal to $5 * x$ elements).

4 Experiments and Results

This section presents the experiments on different countries to measure if, and to which extent, price steering is applied in relation to the Gross Domestic Product of the location from which the user searches on Google Shopping, US version.

4.1 Cities in Different Countries

Two countries feature a significantly different Gross Domestic Product with respect to the third one:

Philippines GDP per capita 7,846.463;
India GDP per capita equal to 6,664.020;
US GDP per capita equal to 57,765.512.

All the values come from the International Monetary Fund website[8] and refer to 2016; they represent the estimated “Gross domestic product based on purchasing-power-parity (PPP) per capita GDP”. For all the experiments, we use English keywords. The three countries have English speaking population. Indeed English is the official language in US, and one of the official languages in India and Philippines. Within these countries, a further selection is on three relevant cities.

Before choosing a US city, we ran experiments investigating price steering with the following targets: San Francisco, Seattle, New York, Los Angeles, Chicago and Miami. We did not unveil significant differences in terms of NDCG. Thus, we opted for choosing New York.

We acknowledge that developing countries, like India and Philippines, feature metropolitan and rural areas, which differentiate between each other for the

[8] https://www.imf.org/external/pubs/ft/weo/2014/02/weodata/index.aspx.

degree of richness[9]. Thus, we consider Manila and New Delhi, since, as metropolis, it should be more likely that people living there belong to the richest part of the population and use to shop online.

Summarizing, we emulate users searching over Google Shopping from Manila, New Delhi and New York.

For the keywords, the lists of product categories are extracted from Amazon.com (due to its richer categorisation, when compared with categories from Google Shopping). In details, 130 terms are considered, belonging to various product categories, including body care, apparel, bags, shoes, car accessories, house accessories (like lightning devices and home appliances), and electronics devices (such as personal audio devices and computer-related products). For each product category, the selection is on common nouns of products (such as luggages, telephones, books), while discarding specific brands.

The experiment settings are as follows:

- Number of locations: 3 (New York, New Delhi, Manila);
- Number of keywords: 130;
- Number of browser profiles for each location: 5;
- Web browser: Firefox 45.0;
- OS: Ubuntu 14.04.

Each browser profile contains its full web history and cookies. Moreover, each profile is kept isolated from the other profiles and it is deleted right after finishing the search. By design, all browsers work simultaneously and send the same query to Google Shopping US.

Each user visit lasts 15 s and the interval time between two searches is a random number between 15 and 30 s. For each country, this work collects the prices associated with the items in the first page of results, for all the 130 keywords, and for all the profiles searching from that country.

To measure the presence of possible price steering, this research considers the NDCG metric, as described in Sect. 3. This requires to compute an ideal list composed of all the results shown to the users, sorted from the most to the least expensive. Thus, for each country, the first 10 products shown to each of the different users are considered, and, among them, the selection is on the 10 most costly ones. The amount of 10 items has been chosen, since, through the experiments, a fact was that users searching from the same city have been shown very similar results. Being basically the difference only in the order, the list of distinct elements tends to be short. The same is done with 3 items only. In fact, statistics report that the top three results on Google account for more than 60% of the average traffic[10]. Thus, even focusing only on the first three results could provide significative outcomes. The two ideal lists are then used to compute, respectively, the measure for NDCG@10 and NDCG@3, per single profile. For each keyword search issued by the users (i.e., the profiles) from one location,

[9] https://data.worldbank.org/products/wdi-maps.
[10] https://chitika.com/google-positioning-value.

the average NDCG@3 and NDCG@10 are computed. These are the NDCG vectors per location. Table 1 reports a snapshot of the most interesting results that we have obtained.

Table 1. An excerpt of average NDCG per location, per sample product

Product	NDCG@3			NDCG@10		
	Philippines	India	US	Philippines	India	US
Boot	0.826	0.85	0.18	0.746	0.755	0.252
Ceiling fan	0.524	0.527	0.245	0.736	0.692	0.475
Desktop	0.114	0.062	0.813	0.219	0.178	0.783
Dress	0.628	0.905	0.282	0.678	0.778	0.371
Helmet	0.592	0.786	0.639	0.639	0.812	0.654
LightScribe	0.138	0.387	0.174	0.27	0.588	0.432
Moisturizer	0.775	0.524	0.48	0.755	0.549	0.523
MP3 player	0.164	0.111	0.165	0.305	0.289	0.312
Pant	0.228	0.271	0.291	0.305	0.446	0.405
Pocket video camera	0.019	0.079	0.574	0.91	0.229	0.28
Portable CD player	0.129	0.097	0.163	0.342	0.352	0.475
Portable DVD player	0.338	0.318	0.859	0.402	0.426	0.87
Television	0.222	0.282	0.542	0.422	0.461	0.592
Universal remote	0.554	0.161	0.597	0.506	0.194	0596
Wheel	0.177	0.202	0.715	0.281	0.311	0.793

Due to connection errors, one of the Philippine profiles had no associated results. Also, for Philippines, a few keywords did not lead to any results: videocassette recorders, totes, umbrellas. Similarly, for US, no results were for totes and umbrellas. This could be due to either a connection error or because Google Shopping provided no results for such items, at time of search. In the case of India, for each keyword, the list of results was never empty.

For Philippines, the top three highest NDCG@3 per location are those related to *briefcase* (0.863), *boot* (0.826) and *blankets* (0.783); For India, *dress* (0.905), *briefcase* (0.88) and *boot* (0.85). For US, *portable DVD player* get 0.859, *desktop* 0.813, and *blankets* 0.8.

Table 2. Average NDCG@10 per location, per sample product category

Category	Philippines	India	US	Category	Philippines	India	US
Apparel	0.389	**0.4**	0.31	Electronic devices	0.195	0.202	**0.4**
Body products	**0.48**	0.455	0.446	All	0.397	0.406	**0.42**

Regarding NDCG@10, per location, for Philippines we obtain *moisturizer* 0.755, *boot* 0.746 and *ceiling fan* 0.736. For India, *fuel system* 0.830, *helmet* 0.812 and *dress* 0.778. For US, *portable DVD player* 0.870, *wheel* 0.793, *desktop* 0.783. Each searched product, in each country, features average NDCG values < 1: in all cases, the search results are not ordered from the most to the least expensive one.

Figure 1 reports mean and standard deviation of the NDCG@3 values per each profile, separately. Similarly, Fig. 2 is for NDCG@10. Overall, from US we obtain the highest results in terms of mean, both for NDCG@3 and NDCG@10, while Philippines feature the lowest ones.

Since keywords can be grouped by categories, we also compute the average NDCG per location, per categories. Table 2 reports the results for the categories Apparel (29 keywords), Electronic Devices (31 keywords), and Body Products (10 keywords). The table shows that the category of products matters. Overall, the three countries have similar average values (last line in the table). Interestingly, for electronic devices US has a much higher value than the other two countries; India features a greater value, compared to US, for apparel; Philippines gets a slightly higher value for body products, *wrt* the other two countries.

Till now, we concentrate on the NDCG metric. It is worth noting that the results on NDCG show how the products prices are ordered, in the different countries, with respect to the ideal list, which, however, is different country per country. Thus, we obtain relative, and not absolute, results: the country which is subject to a higher price steering practice, without however detailing a comparison among the countries (due, in fact, to the difference of each ideal list). In order to give the flavour of the differences in prices in the various countries, we

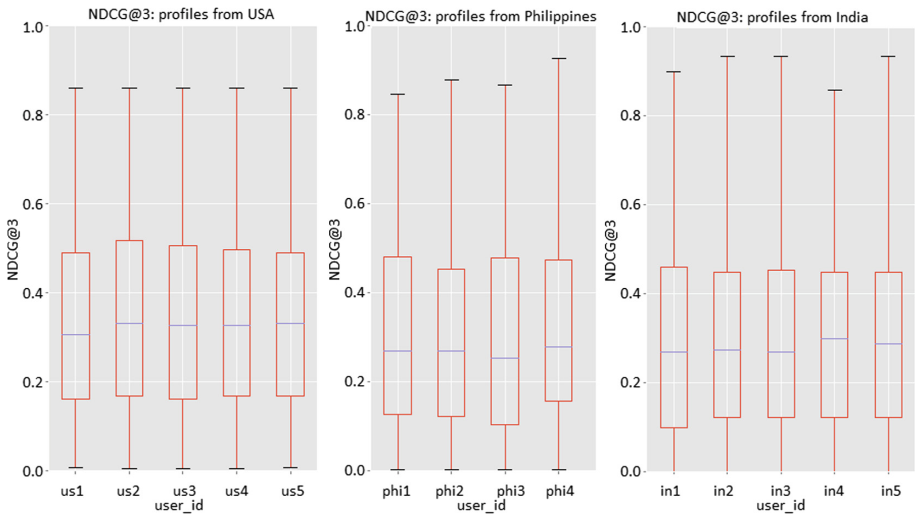

Fig. 1. NDCG@3 per single profile, per location

give some results also in terms of average prices. These are reported in Table 4, where we report average prices for all the product categories, and for sample categories (electronic devices, apparel and body products), considering, at most, the top 10 results per search, and Table 3, which gives an excerpt of the results obtained for the single keywords. Furthermore, to give the flavour of the relative difference among the prices distributions country by country, we compute the coefficient of variation, i.e., the ratio between the standard deviation and the average. This value is reported in Table 4.

Looking at a specific sample category, while Table 2 showed that the most expensive electronic products are shown first in US (with NDCG@10 (per location, per sample category) = 0.4), the average prices are lower than those for India and Philippines (see Table 4). When looking at the coefficient of variation, the relative standard deviation between the products prices is lower. Similarly,

Table 3. Average prices per country and sample products considering the top 10 results

Product	Average prices (US $)			Product	Average prices (US $)		
	Philippines	India	US	Product	Philippines	India	US
Boot	136	136	122	Pant	36	51	73
Ceiling fan	360	340	211	Pocket video camera	151	117	73
Desktop	717	887	368	Portable CD player	50	52	72
Dress	110	100	86	Portable DVD player	66	57	64
Helmet	299	378	315	Television	455	519	625
LightScribe	67	65	128	Universal remote	41	43	95
Moisturizer	22	18	14	Wheel	262	238	178

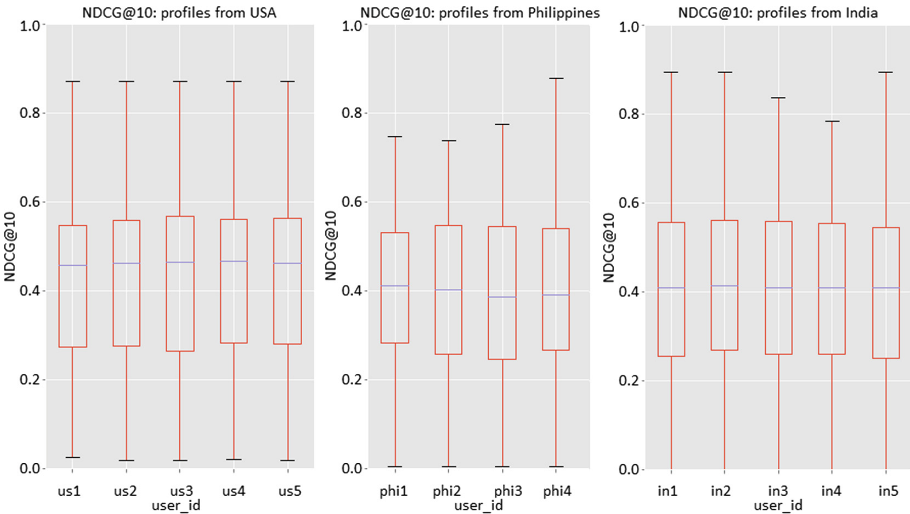

Fig. 2. NDCG@10 per single profile, per location

Table 4. Average prices and coefficient of variation per country and category considering the top 10 results

Category	Average prices (US $)			Coefficient of variation		
	Philippines	India	US	Philippines	India	US
Apparel	62	67	59	1.03	1.08	0.84
Body products	62	67	59	1.03	1.08	0.84
Electronic devices	281	289	149	1.10	1.08	0.93
All	175	173	224	1.4	1.4	5.4

for the specific products, we can notice that the average prices for 7 out of 14 sample products shown in Table 3 are lower searching from US than from the other two countries under investigation (see, e.g., the results for "desktop", with $717, $887 and $368, respectively searching from Philippines, India and US).

As a final remark, we notice that we grounded the experiments on a keyword categorization coming from the Amazon terminology. However, other kinds of choices are possible. An interesting direction is to distinguish durable and not durable goods [10]. In fact, research shows differences buying aptitudes with respect to the two category of goods, i.e., when looking for a durable good, as a laptop, users tend to listen to external recommendations, e.g., from friends. Instead, for not durable goods, they are more influenced by the first results that appear on search engines.

4.2 Statistical Significance of the Experiments

To give statistical significance to the experiments, this work runs permutation tests [20], following the approach proposed in [9,28]. A permutation test on a set of data (in our case, the NDCG values of the location) provides a value, the *p-value*, that represents the probability that a so called null hypothesis is true. Here, the choice of this test is mainly due because it does not require any assumption on the input data.

Considering different countries (Sect. 4.1), the first null hypothesis is defined as follows: the obtained prices of all the investigated products for country x are not distinguishable from the obtained prices for country y. The second null hypothesis is similar, considering however the distinguished categories of products.

Table 5 reports the results in terms of the p-values. Looking at the first two lines of the table (those with p-value equal to 0.050), which refer to all the product categories under investigation, the outcome of the statistical test is that the probability of distinguishing which prices are from Philippines and which are from US is 0.95, with a false rate of 0.05. Even if the p-values of Philippines *vs* US and India *vs* US are lower when considering all the product categories, a pretty good statistical evidence is obtained also for some specific categories (apparel, electronic devices, and "for house").

Table 5. p-values obtained from permutation tests over NDCG values.

Location	Category	p-value	Category	p-value
Philippines vs US	All	**0.050**	Electronic devices	**0.013**
India vs US	All	**0.050**	Electronic devices	**0.056**
Philippines vs India	All	0.89	Electronic devices	0.90
Philippines vs US	Accessories	0.669	For house	**0.096**
India vs US	Accessories	0.525	For house	**0.193**
Philippines vs India	Accessories	0.819	For house	0.740
Philippines vs US	Apparel	**0.136**	Others	0.449
India vs US	Apparel	**0.056**	Others	0.351
Philippines vs India	Apparel	0.629	Others	0.601
Philippines vs US	Body care	0.394		
India vs US	Body care	0.609		
Philippines vs India	Body care	0.711		

5 Conclusions

This paper investigated the impact of locations on the order of price results, searching from common products over Google Shopping US. Differently from previous work in the area, here geographical locations are combined with one of the indicators of the economic performance of the locations, i.e., the Gross Domestic Product. As locations, this work considered cities in three different countries. The analyses aimed at investigating order and averages of prices shown to users. Regarding the order, the considered metric is the difference between the list of price results, as obtained by considering the results of our queries, and an ideal list of results, defined as an ordered list, where the products with the highest prices are first shown to the user. The analysis also considers, at a glance, how the average prices for different categories of products change, location by location. While able to testify the existence of price steering and quantifying its level, country per country, the results of the investigations lead also to unexpected results: even if the experiments on the order of prices highlight the specific country that mostly adheres to the ideal list of results (from the most to the least expensive one), often the average price of the shown results for that specific country is lower than that for the other two countries under investigation. The significance of the obtained results were evaluated by running permutation tests. While satisfactory for certain product categories, they not always succeeded in proving the statistical significance of the experiments. This calls for further investigations. To the best of the authors' knowledge, this is the first work that automatically analyses price steering with respect to GDP values. This introduces a novel, and alternative, methodology, to automatically collect price results exploiting their users' searches on e-commerce search engines. Finally, this work relies on a relatively small set of synthetic accounts and investigated locations. As future work, a natural follow up is to run a wider experimental campaign, considering different e-commerce platforms, more accounts and more locations (like, e.g., different areas in the same city, and different cities in the same country). As a plus, the illustrated methodology can be easily applied on a large scale too.

Acknowledgements. Partly supported by the EU H2020 Program, grant agreement #675320 (NECS: *European Network of Excellence in Cybersecurity*); by the Starting Grants Project DAKKAR (DAta benchmarK for Keyword-based Access and Retrieval), University of Padua, Italy and Fondazione Cariparo, Padua, Italy.

References

1. Carrascosa, J.M., Mikians, J., Rumín, R.C., Erramilli, V., Laoutaris, N.: I always feel like somebody's watching me: measuring online behavioural advertising. In: Emerging Networking Experiments and Technologies (2015)
2. Chen, L., Mislove, A., Wilson, C.: An empirical analysis of algorithmic pricing on amazon marketplace. In: Proceedings of the 25th International Conference on World Wide Web WWW 2016, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland (2016)
3. CNN - International Edition: Amazon seller lists book at $23,698,655.93 plus shipping (2011). http://edition.cnn.com/2011/TECH/web/04/25/amazon.price.algorithm/
4. Conti, M., Cozza, V., Petrocchi, M., Spognardi, A.: TRAP: using targeted ads to unveil Google personal profiles. In: Information Forensics and Security. IEEE (2015)
5. Cozza, V., Hoang, V.T., Petrocchi, M.: Google web searches and Wikipedia results: a measurement study. In: Italian Information Retrieval. CEUR Workshop Proceedings (2016)
6. Cozza, V., Hoang, V.T., Petrocchi, M., De Nicola, R.: Online user behavioural modeling with applications to price steering. In: FINREC 2016 CEUR Workshop Proceedings (2016)
7. Cozza, V., Hoang, V.T., Petrocchi, M., Spognardi, A.: Experimental measures of news personalization in google news. In: Casteleyn, S., Dolog, P., Pautasso, C. (eds.) ICWE 2016. LNCS, vol. 9881, pp. 93–104. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46963-8_8
8. Datta, A., Tschantz, M.C., Datta, A.: Automated experiments on ad privacy settings: a tale of opacity, choice, and discrimination. In: Privacy Enhancing Technologies, vol. 1 (2015)
9. Englehardt, S., Narayanan, A.: Online tracking: a 1-million-site measurement and analysis. In: Computer and Communications Security. ACM (2016)
10. Grandinetti, R.: Concetti e strumenti di marketing. Marketing e vendite, Etas (2002)
11. Granka, L.A., Joachims, T., Gay, G.: Eye-tracking analysis of user behavior in www search. In: Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval SIGIR 2004. ACM, New York (2004)
12. Hannak, A., et al.: Measuring price discrimination and steering on e-commerce web sites. In: Internet Measurement Conference. ACM (2014)

13. Hidayat, A.: PhantomJS (2016). http://phantomjs.org
14. Hupperich, T., Tatang, D., Wilkop, N., Holz, T.: An empirical study on online price differentiation. In: Proceedings of the Eighth ACM Conference on Data and Application Security and Privacy CODASPY 2018. ACM, New York (2018)
15. Jarvelin, K., Kekalainen, J.: Cumulated gain-based evaluation of IR techniques. Trans. Inf. Syst. **20**(4), 422–446 (2002)
16. Kliman-Silver, C., Hannak, A., Lazer, D., Wilson, C., Mislove, A.: Location, location, location: the impact of geolocation on web search personalization. In: Internet Measurement Conference. ACM (2015)
17. Larson, J., Mattu, S., Angwin, J.: Unintended consequences of geographic targeting. Technology Science (2015). https://techscience.org/a/2015090103/
18. Mikians, J., Gyarmati, L., Erramilli, V., Laoutaris, N.: Detecting price and search discrimination on the Internet. In: Hot Topics in Networks. ACM (2012)
19. Mikians, J., Gyarmati, L., Erramilli, V., Laoutaris, N.: Crowd-assisted search for price discrimination in e-commerce: First results. CoNEXT (2013)
20. Nichols, T.E., Holmes, A.P.: Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum. Brain Mapp. **15**(1), 1–25 (2002)
21. Odlyzko, A.: Privacy, economics, and price discrimination on the Internet. In: Electronic Commerce. ACM (2003)
22. Pariser, E.: The Filter Bubble: What the Internet Is Hiding from You. The Penguin Group, London (2011)
23. Pasi, G., et al.: Evaluation of personalised information retrieval at CLEF 2018 (PIR-CLEF). In: Bellot, P., et al. (eds.) CLEF 2018. LNCS, vol. 11018, pp. 335–342. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-98932-7_29
24. Pitoura, E., et al.: On measuring bias in online information. SIGMOD Rec. **46**(4), 16–21 (2018). https://doi.org/10.1145/3186549.3186553
25. Ross, P.: Just How Big Is the eCommerce Market? (2015). https://blog.lemonstand.com/just-how-big-is-the-ecommerce-market-youll-never-guess/
26. Statista: Statistics and facts about global e-commerce (2018). https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/
27. The Wall Street Journal: On Orbitz, Mac users steered to pricier hotels (2012). https://www.wsj.com/articles/SB10001424052702304458604577488822667325882
28. Tschantz, M.C., Datta, A., Datta, A., Wing, J.M.: A methodology for information flow experiments. In: Computer Security Foundations Symposium. IEEE (2015)
29. U.S. Department of Justice: Former e-commerce executive charged with price fixing in the antitrust division's first online marketplace prosecution (2015). https://www.justice.gov/opa/pr/former-e-commerce-executive-charged-price-fixing-antitrust-divisions-first-online-marketplace
30. Vafa, K., Haigh, C., Leung, A., Yonack, N.: Price discrimination in the Princeton review's online SAT tutoring service (2015). https://techscience.org/a/2015090102/
31. Vissers, T., Nikiforakis, N., Bielova, N., Joosen, W.: Crying wolf? on the price discrimination of online airline tickets. In: Hot Topics in Privacy Enhancing Technologies (2014)
32. Yu, S., Hudders, L., Cauberghe, V.: Targeting the luxury consumer: a vice or virtue? a cross-cultural comparison of the effectiveness of behaviorally targeted ads. J. Fashion Mark. Manage. Int. J. **21**(2), 187–205 (2017)