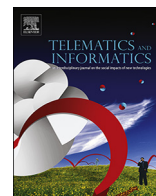




Contents lists available at ScienceDirect

Telematics and Informatics

journal homepage: www.elsevier.com/locate/tele

Challenging Google Search filter bubbles in social and political information: Disconforming evidence from a digital methods case study

Cédric Courtois*, Laura Slechten, Lennert Coenen

School for Mass Communication Research, KU Leuven, Belgium

ARTICLE INFO

Keywords:

Google Search
Filter bubbles
Digital methods
User research
News media

ABSTRACT

This article engages in the debate on supposed online ‘filter bubbles’ by analysing a panel of Google users’ search results on a standardized set of socio-politically themed search queries. In general, the query results appear to be dominated by mainstream media sources, followed at a large distance by civil society and government resources. By means of mixed model regression analyses, with the prominence of different source types in the search results as dependent variables, it was tested whether search results vary across Google Search users. The results indicate that the inclusion of participants as a random effect does not explain variance when controlling for the different query keywords and the time at which the queries were ran. Hence, this study does not support the occurrence of ‘filter bubbles’ in Google Search results in the context of social and political information.

1. Introduction

The overwhelming availability of online information has pushed web developers into designing mechanisms that are able to cope with adverse consequences such as information overload and information anxiety (Bawden and Robinson, 2009). Search engines and especially filtering algorithms work to relieve users from these problems by selecting and prioritizing information into personally tailored selections of relevant information (Bozdag, 2013). Filtering and recommendation systems specifically aim to improve the ease of using platforms by suggesting appropriate and relevant information to individual users (Knijnenburg et al., 2012). This personalization process is usually based on the automated analysis of a broad array of personal data, which allows inferring user preferences, contexts, and interests. Thus, the algorithms exercise power in deciding the types of information that are prioritized over others, whether information is classified favourably, how it is associated with other kinds of information, and how it is eventually filtered and presented to users (Diakopoulos, 2015). It has been argued that through these characteristics, algorithms assume the qualities of a social institution (Napoli, 2014). More specifically, they are characterized by regulative, normative, and cultural-cognitive components that enable and constrain the flow of information (Christin, 2016). As such, they play a pivotal role in the construction of reality as tailors of frames of reference for their users (Just and Latzer, 2016). However, users of search algorithms are generally not transparently informed on how their data are processed and are urged to keep sharing information in order to provide an optimal user experience (Peacock, 2014). Due to the algorithms’ proprietary nature, users are mostly unaware of how algorithms work, causing them to treat these mechanisms as unproblematic means to an end (Gillespie, 2014).

* Corresponding author.

E-mail addresses: cedric.courtois@kuleuven.be (C. Courtois), laura.slechten@kuleuven.be (L. Slechten), lennert.coenen@kuleuven.be (L. Coenen).

<https://doi.org/10.1016/j.tele.2018.07.004>

Received 7 February 2018; Received in revised form 27 May 2018; Accepted 4 July 2018
0736-5853/ © 2018 Elsevier Ltd. All rights reserved.

In light of these considerations, the potential consequences of covert online information filtering has become the subject of concern and fierce debate, with some claiming that it is an indispensable factor in creating bubbles that narrow the scope of information web users come across (Pariser, 2011). It is argued that such constraints have a profound impact by guiding users' belief structures and the actions that stem from it. Currently, there is little rigorous empirical support for any of these claims, which is partly due to the complex nature of studying these phenomena (Ørmen, 2016). This study therefore aims to provide an empirical basis for the debate by systematically documenting and analysing the Google Search results of a panel of users, assessing (a) what information is prioritized, (b) whether individual differences in search results exist, and (c) whether such personalization occurs along the lines of users' ideological positions. The choice for Google is incited by the prominence of online search in acquiring novel knowledge and verifying prior information. It is important to emphasize that the focus of this study is primarily on the overall types of results, and the potential individual variance in Google Search results that is provoked by standardized search queries. Of course, in a naturalistic setting, Google users compose and refine their own queries, while assessing the results. However, to shed light on how Google treats search queries, a strictly controlled environment is a necessity. Still, rather than relying on dummy accounts, the standardized queries are run through actual users' accounts, as if they would search for them themselves. In the following sections, we first address the conceptual background and history of personalized search, as well as its potential vulnerabilities and consequences in order to set the scene for the empirical study.

1.1. Online search environments

Early online search engines analysed web content based on keywords and metatags (Seymour et al., 2011). Such engines faced various challenges, including coping with the massive growth and scale of online information, its dynamic and self-organising character, and the complex hyperlinked structure of the web. These inherent features of the web complicated early search engines' scalability and made them particularly vulnerable for spamming efforts that artificially boost the rankings of third party websites (Langville and Meyer, 2006). This situation changed profoundly by the end of the 1990s with the development of link analysis systems, of which Google Search is undoubtedly the most refined and successful business application up to date. More specifically, Brin and Page (2012) developed the PageRank algorithm at the University of Stanford within the context of a publicly-privately funded research project. PageRank's strength lies in appropriating the networked structure of the web. Its mathematics is based upon the straightforward principle that 'a webpage is important if it is pointed to by other important pages' (Langville and Meyer, 2006, p. 26). Beyond this principle, the algorithm has since then been incessantly refined and supplemented by others, increasing its performance and resilience to spamming and large-scale manipulation (Bar-Ilan, 2007; Google, 2016).

These continuous developments occur within the context of a rapidly expanding media and technology conglomerate. What initially started as a non-commercial research project gradually developed into a promising commercial venture (Steiber and Alänge, 2013). Since its inception, Google has grown from a search engine provider into the diversified publicly traded enterprise Alphabet, catering the needs of over a billion users through a proliferation of adjacent services (Vaidhyanathan, 2012). Through this process, the development of a viable business model became an evident necessity (Finkle, 2012). Google now essentially harnesses refined user data from its various popular services to target the online advertising market. It lucratively mediates between advertisers and web platforms by enabling the former to individually target search results to specific consumer profiles, while also providing advertising revenues to the latter (Kang and McAllister, 2011; Lee, 2011). Consequently, the company's objective has become to invest in a wide array of integrated services that keep the so-called 'virtuous cycle of big data' running (Harrison, 2015). This cycle implies that Google appropriates user data to continually improve services and its supporting algorithms to create a better experience. This, in turn, attracts and retains more marketable users, which increases the volume and the granularity of collected user data that supports further improvements and commercial micro-targeting efforts. Thus, the structural boundaries of the algorithms' design are inevitably informed by a clear set of values that are inevitably market-oriented rather than inspired by fairness and representativeness of information (Van Couvering, 2007).

It is estimated that Google annually changes its search algorithm about 500 to 600 times, which usually comprises minor adjustments (Moz, 2016). In 2009, however, Google introduced a pivotal novelty that formed the basis for search engine personalization (Horling, 2009). To enhance the search engine user experience, mechanisms were implemented to infer user preferences. This system is based on the intuitive logic that the relevance of information varies between users and their contexts. Hence, it infers and accounts for the variable meaning of the same keywords for different users. Divergent types of information sources were programmed into the algorithm: on the short term, recent search queries inform about context, whereas over time refined user profiles are built to get a grasp on long-term patterns of individual preferences and characteristics (Smyth et al., 2011).

Research on the occurrence of personalization in Google products has rendered mixed results. One of the first studies that focused on Google Search involved running a standardized set of 80 queries on actual user accounts, comparing its results with the results of the same queries run without logging in (Hannak et al., 2013). The results were then analysed for overlap (i.e., Jaccard Index) and differing rank orders (i.e. Kendall's Tau). The results indicated that, on average, about twelve per cent of searches were personalized, peaking for queries on news and political issues. A follow-up study, centred on the factor of location, showed that up to 34% of search results varied in terms of the results that were shown (Kliman-Silver et al., 2015). Apart from that, also substantial variations in rank orders of the search results were found. Nevertheless, personalisation does not seem to emerge in a straightforward manner. Hoang et al. (2015) experimented with training Google Search accounts, rendering differences in up to one quarter of the returned results. However, the outcomes of the training did not appear to directly correspond with the contents of the account training. Still, this evidence is somewhat tentative as it dates back several years, while the service further developed. Moreover, the majority of studies were based on analysing mock-up accounts. On the other hand, a variety of studies, particularly focused on Google News repeatedly

Download English Version:

<https://daneshyari.com/en/article/10225844>

Download Persian Version:

<https://daneshyari.com/article/10225844>

[Daneshyari.com](https://daneshyari.com)