Detecting and visualizing filter bubbles in Google and Bing

Tawanna R. Dillahunt

School of Information, University of Michigan

Michigan, MI, 48109, USA

tdillahu@umich.edu

Christopher A. Brooks

School of Information, University of Michigan

Michigan, MI, 48109, USA

brooksch@umich.edu

Samarth Gulati

School of Information, University of Michigan

Michigan, MI, 48109, USA

samarthg@umich.edu

Abstract

Despite the pervasiveness of search engines, most users know little about the implications of search engine algorithms and are unaware of how they work. People using web search engines assume that search results are unbiased and neutral. Filter bubbles, or personalized results, could lead to polarizing effects across populations, which could create divisions in society. This preliminary work explores whether the filter bubble can be measured and described and is an initial investigation towards the larger goal of identifying how non-search experts might understand how the filter bubble impacts their search results.

Author Keywords

Search; Filter bubble; Google; Bing; Personalization; Stereotype threat

ACM Classification Keywords

H.5.m. Information interfaces and presentation Misc.

Introduction

People from all over the world use search engines to access information, services, entertainment and other resources on the web. Built on revenue largely derived from advertisements, web search is a multi-billion dollar industry. Many search engines, including the top 2 most popular search engines Google and Bing [16],

customize user search results based on user's search history, location, and past click behavior. While this generates personalized results, this leads to an effect known as the "filter bubble", a term coined by Eli Pariser [12] to describe how recommendation engines protect people from certain facets of the world. Pariser used a Google search for "BP" across two people as an example. One searcher received links to the Deepwater Horizon oil spill and the other received links to investment news – two very different perspectives from the same search query.

In this study, we investigate whether the filter bubble can be *measured* and *described* in a way that might be understood by non-search experts.

Related Work
The underlying goal of most research around search has been to understand the search behavior of individuals to improve system performance of search engines, user satisfaction, and to inform the design of better search engines [7].

Users know little about search, search engines and their impact
Despite the pervasiveness of search engines, most users know little about the implications of search engine algorithms and are unaware of how they work [3]. Past research finds that search engine users do not know the difference between paid and unpaid search results [1]. In fact, people using web search engines assume that search results from search engines are unbiased and neutral [4].

Information literacy studies focus on how ordinary users have varying search abilities and convey that

many users are not well prepared to deal with the bias of search [6]. A minority of searchers use complex search. Most individuals use short search terms and seldom click beyond the first page of results. In fact, most people believe that the best results appear at the top of the results even when results are intentionally scrambled [8].

Given what is known about how demographic and geographical factors influence search results, it is unclear how this interacts with search engine bias. In addition, understanding how the filter bubble works is technical, nuanced, and often a web search company's corporate secret.

Private Browsing and Deterring the Filter Bubble
Private browsing mode is a relatively new feature of modern web browsers. In private browsing mode, some elements of user tracking, such as cookies and user history, are deleted after every session. While this does not render the user completely invisible from tracking, it can reduce the amount of tracking data a website can collect, and limit the amount of personalization ("bubbling") a service provides. Despite enabling this mode, web services make use of the user-agent string to access user information such as their browser, language, and location (e.g., IP Geolocation services). This enables web browsers to build personalized profiles, which may stereotype the user and lead to personalized results.

While we can aim to reverse engineer the filter bubble, it is constantly changing as corporations modify and improve upon their algorithms [12]. Instead, we take a different approach and explore whether the filter bubble can be *measured* and *described* in a way that might be

Figure 1 - Screenshot of Amazon Mechanical Turk (AMT) Link (left) and selected Google and Bing Search Queries (right). We selected different search queries representing high stakes information seeking needs where we felt the filter bubble may be present. Timeline: Oct. 28th to Nov 13th 2014.

understood by non-search experts. Our work aims to extend prior work to combat filter bubbles [10].

Methodology

To measure the filter bubble, we recruited 20 users from Amazon's Mechanical Turk to conduct five unique search queries. Users were instructed to: 1) copy a search link to their clipboard, 2) open a private browsing window, 3) paste a link in the address bar, and 4) press enter. Instructions were provided on how to open a private browsing window for the web browser the participant was using. As a final step, we asked participants to copy the HTML results from the "view source" option in their browser, and paste the text into a text box in the main study window. Users repeated this process ten times—once for each of the search URLs listed in Figure 1 (right).

Data Analysis

We captured a unique user ID, search query, search result, and user-agent string into a JavaScript Object Notation (JSON) formatted object. Our goal was to analyze the extent of the differences in search results, exclusive of advertisements. Our work is similar to [5]; however, we explored both Google and Bing and search in a somewhat more realistic context.

Search engines often return a different number of results. This could vary depending on user screen real estate, or the speed at which the results are returned. To normalize our dataset, we collected only the first eight results from each search engine, and pruned users who had less than eight results.

To conduct our analysis, we compared lists of user search results across the same search query. We used the Kendall Tau Rank Distance (KTD) [9] to count the number of pairwise disagreements between two ranking lists. This metric identifies how different two sets of results are from one another – a KTD value of 0 represents two perfectly similarly ranked lists, whereas a KTD value of 1 represents two completely dissimilar lists. This metric does not take into account the size of the distance between two items in a list, just whether their rank ordering in the lists are similar or not. We also calculated the level of connectedness of the distances between search results using the clustering coefficient. As every search result was compared to every other search result (e.g. we have a k-map of KTD values), we used a weighted clustering coefficient [13]. A low clustering coefficient (0) for a search result means that, within a given search engine, there is little personalization within the search results, while a high

clustering coefficient (1) indicates a high within-search engine personalization.

We use the overall average and standard deviation to describe the amount of agreement within a group of search results. A low average suggests minimal disagreement for a given search term, while a high average suggests there is significant disagreement.

Results

To aid our readers in understanding how to interpret KTD, Figure 3 shows an example of the results of one search using the Bing search engine. In this image we have weighted the differences between user searches as pairwise KTD, and then visualized the results using the Fruchterman-Reingold [2] force-directed algorithm as provided by the NetworkX library [11]. This method attempts to minimize distances between nodes, which have low KTD distances while maximizing distance between nodes with high KTD distances. A given edge connecting two nodes is thus the KTD between those nodes and the length of this edge is mediated in part by the rest of the network ties so that the network can be shown in two dimensions. The result shows the amount of clustering that exists within the system for high numbers of people. In this figure, we see there are two clear outliers – people (user identifiers are in node labels) with high KTD from others. These individuals are "bubbled" compared to the rest of their peers.

Figure 2 shows two columns of five search results, the first column being from the Bing search engine and the second being from the Google search engine. By comparing across the columns one can see the effect search engine algorithms have on the same query. By comparing down the columns one can see the effect

different queries have on the filter bubble within a given search engine. This allows for cross search engine comparison. For example, the difference in Kendall's Tau Distance values is much larger between asthma and jobs in Google than in Bing. Does this mean that Goggle "turns off" the filter bubble in some searches? If so, is it possible to identify which searches?

Conclusion and Future Work

The aim of this work was to explore whether the filter bubble can be measured and described. The visualizations from Figure 2 demonstrate partial success, and the key contributions of this work.

As noted in the related work, most individuals seldom click beyond the first page of results, and most people believe that the best results appear at the top of the results even when results are intentionally scrambled [8]. Given this finding, it may be appropriate to more heavily weight, for example, a disagreement between two sets of search results A and B where one URL is the first result in A and the fourth in B versus the first result being ranked in A and the second in B. Kendall's Tau Ranking Distance instead gave an approximation of the difference between two sets of results. We will consider replacing KTD with the *expected weighted Hoeffding* distance, another dissimilarity function proven successful in prior work [14]. To normalize our dataset, we only collected the first eight results from each search engine and pruned users who had less than eight. This limitation will later be addressed as this led to dissimilar comparisons across users.

Going forward, we plan to perform user tests to explore alternative visualizations and evaluate whether these visualizations are understandable among non-search

experts. This is a new research topic and no guidelines exist for these types of visualizations. We would like to explore what it would be like for users to receive this type of information while browsing. We also hope to identify which factors are causing the filter bubble in each query. We will evaluate differences in the user-agent string to achieve this goal.

We would also like to understand if some search engines are less likely to exhibit the filter bubble effect. For example, DuckDuckGo (http://dontbubble.us/) advertises an alternative search engine that breaks the filter bubble. Finally, we removed advertisements in this version of the work; however, exploring the effects of the filter bubble in advertisements could be interesting. For example, in an investigation of the delivery of personalized ads for public records doing basic name searches, Sweeney found statistically significant results in the discrimination in ad delivery across two websites based on names alone [15].

References

[1]   Fallows, D. (2005). Search engine users. Pew Internet & American Life Project. Retrieved March 18, 2007 from http://goo.gl/2Veaev.

[2]   Fruchterman. T. E. Reingold. 1991. Graph drawing by force-directed replacement. *Software—Practice and Experience* *21* 1129-1164.

[3]   Gerhart, S. (2004). Do Web search engines suppress controversy? *First Monday,* *9*(1).

[4]   Goldman, E. Search Engine Bias and the Demise of Search Engine Utopianism, 8 Yale J. L. & Tech. 188 (2006).

[5]   Hannak, A., Sapiezynski, P., Kakhki, A. M., Krishnamurthy, B., Lazer, D., Mislove, A., and Wilson, C. Measuring personalization of web search. In WWW, 2013.

[6]   Hargittai, E., & Hsieh, Y. P. (2013). Digital Inequality. In W. H. Dutton (Ed.), The Oxford Handbook of Internet Studies (pp.129-150). Oxford, UK: Oxford University Press.

[7]   Hsieh-Yee, I. (2001). Research on Web search behavior. *Library and Information Science Research,* *23*(2), 167-185.

[8]   Jansen, B.J., Spink, A. (2006). How are we searching the World Wide Web? A comparison of nine search engine transaction logs Information Processing and Management, *42* (1), pp. 248–263.

[9]   Kendall's Tau Distance. (n.d.). In *Wikipedia*. http://en.wikipedia.org/wiki/Kendall_tau_distance.

[10] Munson, S.A. & Resnick, P. (2010). Presenting diverse political opinions: how and how much. pp. 1457-1466.

[11] NetworkX, http://networkx.github.io/.

[12] Pariser, E. (2011). The filter bubble: What the Internet is hiding from you. London: Viking/Penguin Press.

[13] Saramäki, J., Kivelä, M., Onnela , J.P., Kaski, K., Kertesz, J., 2007. Generalizations of the clustering coefficient to weighted complex networks. Physical Review E Stat. Nonlinear Soft Matter Phys. 75 027105. http://jponnela.com/web_documents/a9.pdf

[14] Sun, M., Lebanon, G., and Collins-Thompson, K. "Visualizing differences in web search algorithms using the expected weighted Hoeffding distance," in Proc. 19th Int. Conf. World Wide Web.

[15] Sweeney, L. Discrimination in Online Ad Delivery. Communications of the ACM, 2013, Vol. 56 No. 5, 44-54.

[16] Top 15 most popular search engines | January 2015. *The eBusiness Guide.* *http://www.ebizmba.com/articles/search-engines*.
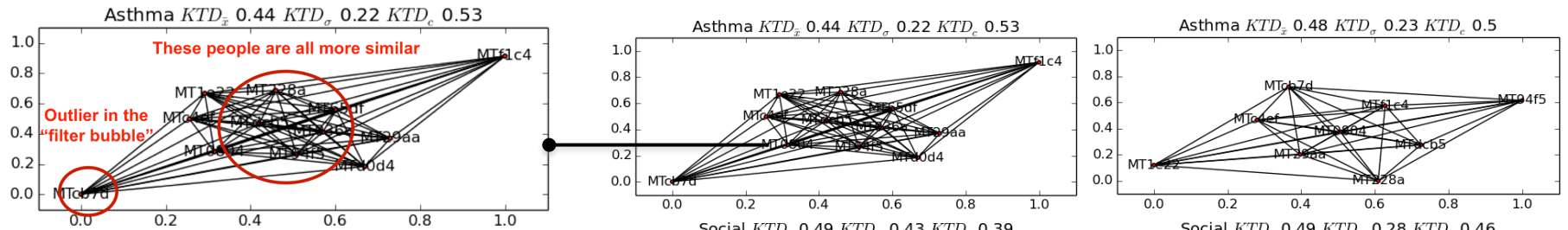
Figure 2 – Bing single search result (Asthma query). Note the cluster of similar search results in the middle, and the two strong outliers. These outliers are in the "filter bubble", compared to their peers.
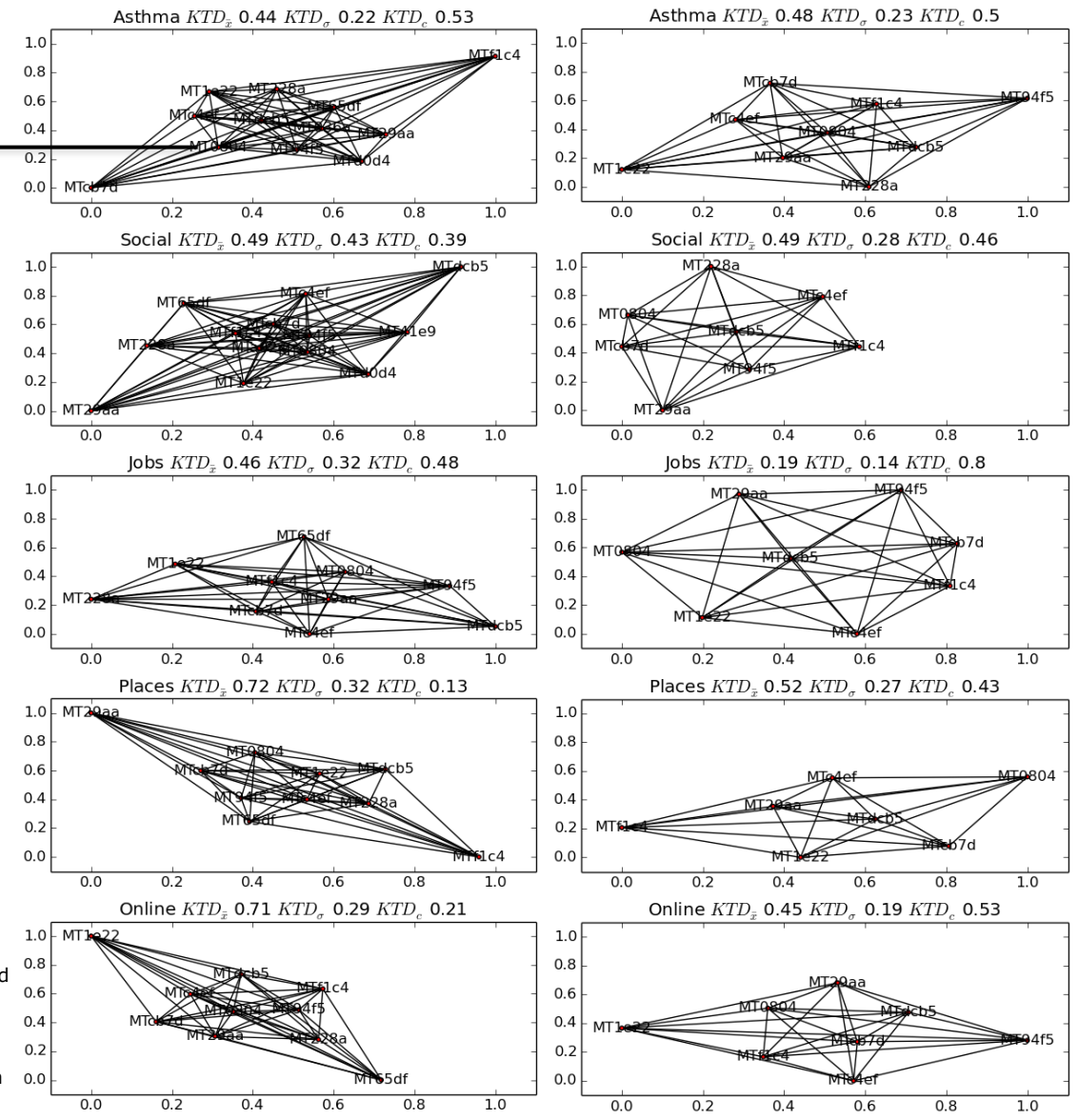
Figure 3 – Bing search results (left column); Google search results (right column). $KTD_{\bar{x}}$ is the mean distance for the graph, and KTD_{σ} is the standard deviation. The clustering coefficient, KTD_c describes the strength of connectedness in the graph. Reading across the columns compares individual query results between the two search engines, while reading down a column compares search results across queries but within a search engine.