# YTREX: crowdsourced analysis of YouTube's recommender system during COVID-19 pandemic

Leonardo Sanna[1], Salvatore Romano[2], Giulia Corona[3], Claudio Agosti[4]

[1] University of Modena and Reggio Emilia, Modena, Italy
leonardo.sanna@unimore.it
[2] University of Padova, Padua, Italy
salvatore.romano.3@studenti.unipd.it
[3] University of Milano, Milan, Italy
giulia.corona1@studenti.unimi.it
[4] Tracking Exposed, https://tracking.exposed/
claudio@tracking.exposed

**Abstract.** Algorithmic personalization is difficult to approach because it entails studying a multitude of different user experiences, with a lot of variables that are difficult to control. Two common biases are frequent in experiments: relying on corporate service API and using synthetic profiles with small regards of regional and individualized profiling and personalization. In this work, we present the result of the first crowdsourced data collections of YouTube's recommended videos via YouTube Tracking Exposed (YTREX). Our tool collects evidence of algorithmic personalization via an HTML parser, anonymizing the users. In our experiment we used a BBC video about COVID-19, taking into account 5 regional BBC channels in 5 different languages and we saved the recommended videos that were shown during each session. Each user watched the first five second of the videos, while the extension captured the recommended videos. For each completed session, we took into account the top-20 recommended videos, looking for evidence of algorithmic personalization. Our results showed that the vast majority of videos were recommended only once in our experiment. Moreover, we collected evidence that there is a significant difference between the videos that we could retrieve using the official API and between what we collected with our extension. These findings show that filter bubbles exist and that they need to be investigated with a crowdsourced approach.

**Keywords:** Independent algorithm analysis, crowdsourced data collections, network analysis, official API, COVID-19, YouTube.

# 1 Introduction

## 1.1 Algorithmic personalization

Algorithmic personalization is by now part of our lives. In fact, we use recommendation systems for a remarkably high number of tasks, ranging from working to free time. Each time we google something for our work, an algorithm is selecting what is most relevant for us, the same happens when we scroll our Facebook feed and when we use our Netflix or Spotify account. We may say that algorithms are the technological solution to the information overload in which we live, daily.

However, the majority of these services are owned by private corporations which use black-box algorithms to curate the content selection for their users. In the last years, these platforms have been at the stake of academic research, in particular for what concerns the so-called “fake news debacle”, with a lot of research focusing on misinformation spreading and on polarization of the public debate[1, 2, 3].

For the sake of clarity, one important distinction has to be made about this. Research that takes into account information spreading, online debate and user engagement is a study on echo chambers. Echo chambers, although not well defined in literature, are social phenomena based on ideological affinity and they are relatively accessible for research. Instead, in this paper we are interested in studying the filter bubble [4], that is the direct effect of algorithmic personalization. To use the words of Zimmer [5], echo chambers are made by users, while filter bubbles are made by algorithms.

At the beginning of 2020 the academic debate started questioning the existence of the filter bubble effect, following the ideas proposed by Bruns [6] that claimed that there is no evidence that echo chambers and filter bubbles exist outside the academic theory. On the other hand, empirical research on algorithmic personalization is still quite fragmented and we believe that this happens because of the lack of a shared methodology among the researches. The fact that we must take into account user experience is problematic because of the number of uncontrollable variables and also because it requires user collaboration or fabricated profiles.

YouTube, from its part, provides information about its own algorithm, but just related to the general structure of the recommended system algorithm [7, 8], but we don’t know exactly how many variables are used for the personalization process and how. We can access just a bunch of metadata related to the algorithm personalization requiring our personal data to the platform with a "data subject access request". The site also provides an official API, often used by researchers but is not clear the relation between those data and real data view with real profiles, as we will see later.

In the following sections we review the latest methods for algorithmic personalization analysis.

## 1.2 Methods for algorithmic personalization analysis.

There are a number of studies that claim that are focused on filter bubbles but the majority of these studies is not actually taking into account algorithmic personalization but, instead, it is inquiring ideological preference with methods of the social sciences [9, 10].

These approaches do not consider the fact that algorithmic personalization is a product of the interaction between users and platforms and it is essentially passive since users have very limited or no control of their personalization. Hence, if we approach the filter bubbles starting from user behavior, we are not studying the filter bubbles but the echo chambers. In our opinion, trying to infer conclusions on algorithmic personalization starting from its alleged effects might produce misleading outcomes.

In this work we are proposing a crowdsourced approach as it was already experimented by Robertson et al. on Google SERP [11]. In their study the authors stated that they found very little evidence of filter bubble effect. Although we found their methodology robust and well-suited for the study of filter bubbles, we believe that their conclusions should be reconsidered taking into account that:

1. algorithmic personalization is platform-specific, as every platform as its own algorithm
2. Google SERP might be not the best platform to inquiry filter bubble in its original meaning of “informational bubble”, as it is produced by an intentional research
3. the “in incognito mode” of the web browser Chrome is not a completely clean navigation, as it keeps a certain level of personalization that is based on geographic location.

Regarding point (1) one of our previous experiments we found evidence of algorithmic personalization on Facebook using synthetic profiles [12] so that we could control all the variables. A previous explorative analysis on the YouTube algorithm also show similar results[1]. Regarding point (3) we must recall that controlling the variable is one of the main difficulties while studying algorithmic personalization, also because we have to proceed by trials and errors, since we do not know exactly which variables influence algorithmic decisions.

We hence propose to investigate specific effects (such as ideological preference) using fabricated profiles and to use crowdsourced approach to gain evidence of existence of personalization, thus, to account for the fragmentation of content distribution.

**Approaching filter bubbles on YouTube.** According to what YouTube (YT) Chief Product Officer Neal Mohan said, 70% of videos watched on YT are recommended videos. This huge proportions make the study of filter bubbles within YouTube particularly relevant. We propose a crowdsourced methodology to investigate user personalization within the recommended videos on YouTube. The context of COVID-19 was an occasion great occasion to test our tool YTREX, looking at how YT distributed its content related to the pandemic. An extensive review on YouTube research has been conducted by Arthurs et al. [13] and they highlighted that the platform is vastly

[1] University of Amsterdam, Digital Methods Summer School. Algorithm exposed Investigating Youtube personalization with YTREX. https://wiki.digitalmethods.net/Dmi/SummerSchool2019AlgorithmsExposed

understudied. To the best of our knowledge other studies on YT filter bubbles that propose a crowdsourced approach do not exist, by now. Related works, use YT API or other methods that are not user-centered, nor they collect empirical data directly from users' browsers [14, 15, 16, 17]

# 2 Tool: YouTube Tracking Exposed

## 2.1 How it works

The browser extension (add-on) of Tracking Exposed[2] collects evidence from the metadata that is observable on the web page when the user lands on the homepage, watches a video, or does research on the YouTube website. It creates cryptographic key pairs to ensure the user can access her/his data. It is necessary because the tool does not have an email address, Google profile, or any other authentication method based on personal data. The tool collects separate contributions for each browser with the add-on installed.

The data are collected in three phases:

1. **Collection**: the add-on takes a copy of the HTML when the browser is watching a video. Four buttons appear on the top left of the screen (Fig.1), when the add-ons is installed and enabled by the popup. The color code represents the different statues.

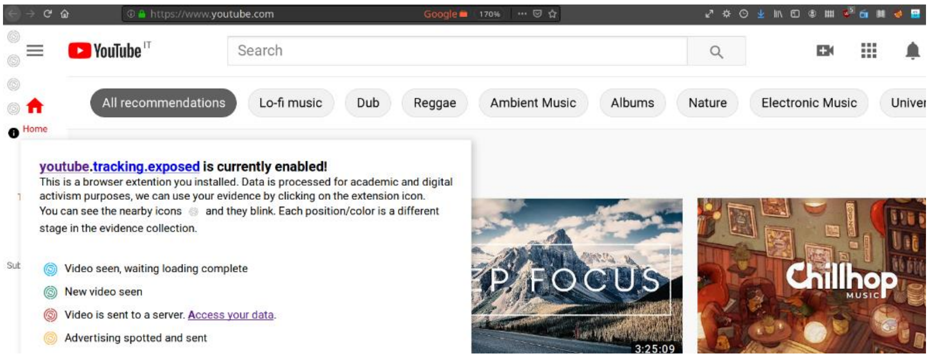


**Fig. 1.** Screenshot of what the browser extension shows while navigating on YouTube.

2. **Parsing**: server side, the HTML is processed and metadata are extracted. The information is then organized in a dataset. In the HTML there are many different data that might be analyzed to extract metadata. We did not yet extract all the possible information, especially we avoid any unique tracker that might become personal data if collected. On the other hand, the YTREX project still has room for improvement, and we might not have yet mapped 100% of the potentially interesting metadata for YouTube algorithm analysis.

---

[2] youtube.tracking.exposed

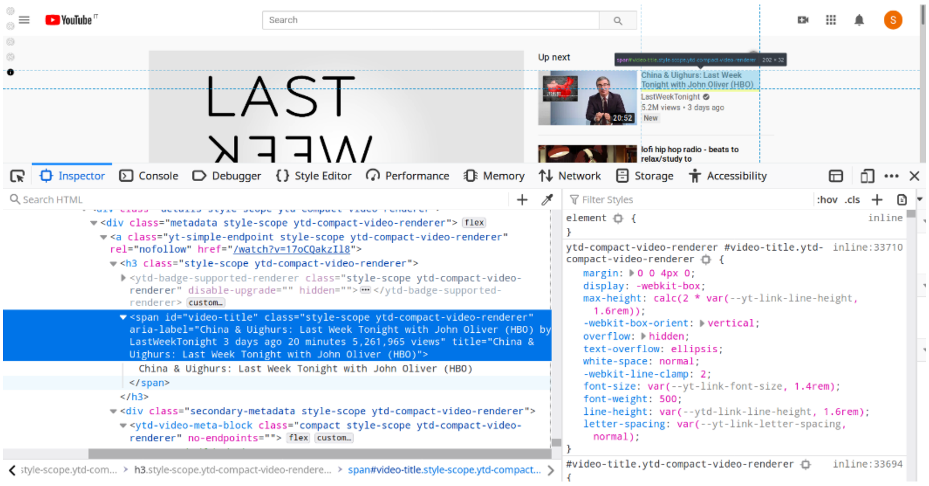


**Fig. 2.** HTML inspection of a recommended video on YouTube and its aria-label

Inspecting the HTML of a recommended video (Fig. 2), you might see the data field named aria-label[3], it is a text field meant for accessibility and contains a compacted, but human formatted, set of information useful for researchers. Because of the localization, YouTube produces aria-label with strings that change accordingly to the user interface Language.

```
CORONAVÍRUS é uma FARSA? Melhor não assistir o vídeo... by Maro Schweder - espiritualidade e autoconhecimento 1 week ago 26 minutes 47,109 views

街拍英國佛系抗疫｜倫敦市中心如死城｜廁紙意粉全城缺貨？ di Mario Hau 孝仔 4 giorni fa 10 minuti e 35 secondi 190.301 visualizzazioni

Frank Cuesta, sobre el origen del COVID-19: “En agosto ya estaba pasando algo” de esRadiovideos il y a 2 semaines 6 minutes et 18 secondes 2 788 191 vues
```

Color code:

* Title
* Video duration, get converted in seconds
* Number of visualizations
* Author name
* Since when the video have been published

3 For reference shttps://developer.mozilla.org/en-US/docs/Web/Accessibility/ARIA/ARIA_Techniques/Using_the_aria-label_attribute

**Fig. 3.** Examples of different aria-label formatting.

This natural language conversion managed by our aria-label parsing library[4], we might externalize it as an independent library, once we figure out how to maintain the list of fixed terms that scale up proportionally to the language supported by YouTube. The sum of session information, video watched, and recommended video, produce the data unit with the format of Fig. 4


```
"evidence": 33,
"login": true,
"id": "c8eecebbdcd9badcafdc",
"savingTime": "2020-03-26T21:33:58.071Z",
"clientTime": "2020-03-27T02:45:01.000Z",
"uxLang": "es",
"recommendedId": "e2b9eb82cb36fbc8445f42a63f9f7b8c93e65292",
"recommendedVideoId": "DSuwaZEF9ZY",
"recommendedAuthor": "BBC News",
"recommendedTitle":  "Coronavirus:  How  bad  is  the  situation  in
Europe? - BBC News",
"recommendedLength": 79860,
"recommendedDisplayL": "22:11",
"recommendedLengthText": "22 minutos",
"recommendedPubTime": "2020-03-21T02:45:01.000Z",
"recommendedRelativeS": 518400,
"recommendedViews": 3435,
"recommendedForYou": false,
"recommendedVerified": true,
"recommendationOrder": 34,
"recommendedKind": "video",
"watchedVideoId": "A2kiXc5XEdU",
"watchedTitle": "How do I know if I have coronavirus? - BBC News",
"watchedAuthor": "BBC News",
"watchedChannel": "/user/bbcnews",
"watchedPubTime": "2020-03-21T00:00:00.000Z",
"watchedViews": 682214,
"watchedLike": 6331,
"watchedDislike": 223,
"hoursOffset": 46,
"experiment": "wetest1",
"pseudonym": "nachos-taco-manicotti",
"top20": false,
"isAPItoo": false,
"step": "English",
"thumbnail":"https://i.ytimg.com/vi/DSuwaZEF9ZY/mqdefault.jpg"
```


**Fig. 4.** Example of a data unit obtained with YTREX

3. **Research and data-sharing**: YTREX was born to support independent analysis and privacy-preserving sharing of the algorithmically-powered circulation of videos. Every video observation has a dynamic number of related videos (if the watcher scrolls the video page down, the browser loads 80 or more related videos, but the default for users who do not scroll down, is to receive and display only the first 20 related videos). Every related video becomes a single row, a data record with its own unique ID. Interconnecting these with *metadataId*, the researcher might re-group all the related videos belonging to the same evidence, as they were displayed

[4] https://github.com/tracking-exposed/yttrex/blob/master/backend/parsers/longlabel.js

to the watcher. Certain fields such as *logged*, *pseudo*, and *savingTime*, are the same across the same id because they depend on the collection condition. *RecommendedVideos*, *recommendedAuthor*, and other recommended- fields, changes in each row according to the related video described; *recommendedId* is generated for each row, and should be used as guarantee of unique field.

According to the definition provided by Sandvig [18], the tool enables the user to potentially four of the five methods of algorithmic audit: Noninvasive User Audit, Scraping Audit, Sock Puppet Audit, if they have the know-how to use bots, and Crowdsourced or Collaborative Audit, as the experiment presented on this paper.

The database collected for this paper is available on Tracking Exposed website[5] and the code is available on GitHub[6] protected by AGPL v3 license.

## 3 Experiment

We made a call for participation on our website to select the participants. Every participant joined the experiment for free and on a voluntary basis. The procedure involved the visualization of five videos about COVID-19 prevention, produced by the BBC channel, one for each language into which the channel is translated: Chinese, Spanish, English, Portuguese, Arabic[7]. We did not provide additional information about the minimum time to spend watching the videos, loading the page is enough to collect the HTML and enabling the parsing of all the information about the first twenty related videos suggested by YouTube. The dataset includes only complete sessions made of seven observations each, throughout a dedicated node.js script.

We were also able to record the users interface language analyzing the timestamps assigned to the watched video by the platform (One year ago, Hace un año...). Participants could choose to perform the test logged with their personal account or without, the tool records if the user is logged or not, without collecting any data related to the specific account.

### 3.1 Official API comparison

The same day of the test, we retrieved via official YouTube's API the related videos for the six videos present in the methodology.

We performed the API request in six languages and with six region codes. 50 videos were retrieved in each API request, all the videos were put together to create a new variable in original csv, declaring if the video suggested during the experiment was

[5] https://youtube.tracking.exposed/data/

[6] https://github.com/tracking-exposed/youtube.tracking.exposed

[7] Chinese: https://www.youtube.com/watch?v=Lo_m_rKReyg,
Spanish: https://www.youtube.com/watch?v=Zh_SVHJGVHw,
English https://www.youtube.com/watch?v=A2kiXc5XEdU,
Portuguese: https://www.youtube.com/watch?v=WEMpIQ30srI,
Arabic: https://www.youtube.com/watch?v=BNdW_6TgxH0

present in the official API or not. This allowed us to compare our crowdsourced output with the data provided by YouTube.

# 4 Findings

## 4.1 Evidence of filter bubbles

The distribution of the recommended videos is clearly skewed as shown in Fig. 5. We investigated the distribution of recommended videos taking into account the language of the starting video, the language of the browser and also considering whether the user was logged or not. No matter of which variable we took into account we always obtained a skewed distribution, as shown in the example of Fig. 6.


**Fig. 5.** Frequency distribution of recommended video in our dataset.

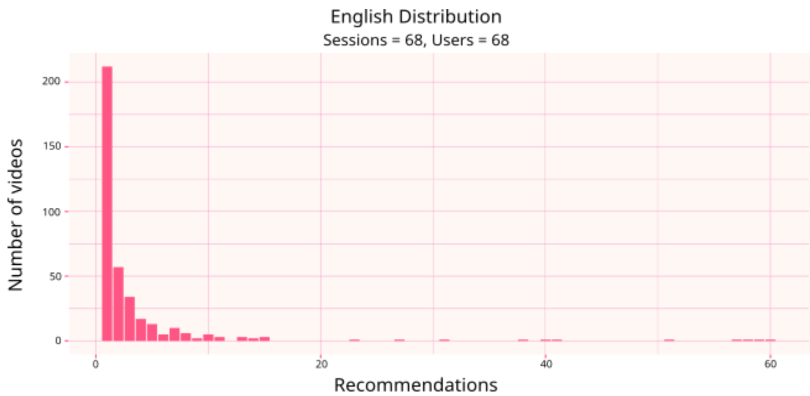


**Fig. 6.** Frequency distribution of recommended videos starting from the BBC video in English.

Our findings show that the vast majority of videos are recommended very few times (1-3 times), regardless of the variable considered. This distribution is significantly positively skewed according to Fisher's skewness coefficient (>2). Summing up, 57%

of the recommended videos have been recommended only once and only around 17% of the videos have been recommended more than 5 times during our experiment.

These results highlight that the filter bubble is real, and that algorithmic personalization produces a high fragmentation of recommended content among YouTube users. Another relevant finding for the study of algorithmic personalization is the huge difference we found using YT API and our tool. For users logged into their Google account only 11% of the recommended videos could be retrieved using the API as showed in the following section, in Fig. 9.

### 4.2 Network analysis

We performed a network analysis using Gephi [19] to better understand and visualize how the recommender system creates a filter bubble around users watching the same video the same day. Thanks to the Medialab's tool *Table2net*[8] we extracted a network file from the csv file. We created a bipartite network linking two types of nodes: users' pseudonyms and suggested video's ID.

In the graphs (Fig: 7, 8, 9) we used a circular layout algorithm [20] to dispose of all the users in a circle. Our aim was to show all the participants in the same positions, pointing in the same direction, because they are actually performing the same task: in the examples they are watching the video from the English version of BBC channel "How do I know if I have coronavirus? - BBC News.". This representation allowed us to show how they are watching the same video, but they are getting a differentiated configuration of suggested videos.

Then, we performed a Force Atlas 2 Algorithm to place each related video close to the users who received that suggestion by the platform. On the one hand this technique highlights the network centrality of a bunch of videos, homogenously suggested across users (the ones in the center of the graph), on the other hand we can clearly see videos suggested to singular users (those who are external to the users' circle, really close to just one pseudonym).

The size of the nodes is based on the degree of each node: in a range between size 15 and size 60, each user and each recommended video is big in relation to the number of links that it has. We can see that some of the participants performed the test twice, because their node is bigger than the others, at the same time the recommended videos in the center of the graph are bigger because they have been suggested more than the others. Beware that, for a graphical compromise, the nodes with a degree minor than 15 have the same shape, likewise the nodes with a degree higher than 60 are all the same.

[8] https://medialab.github.io/table2net/

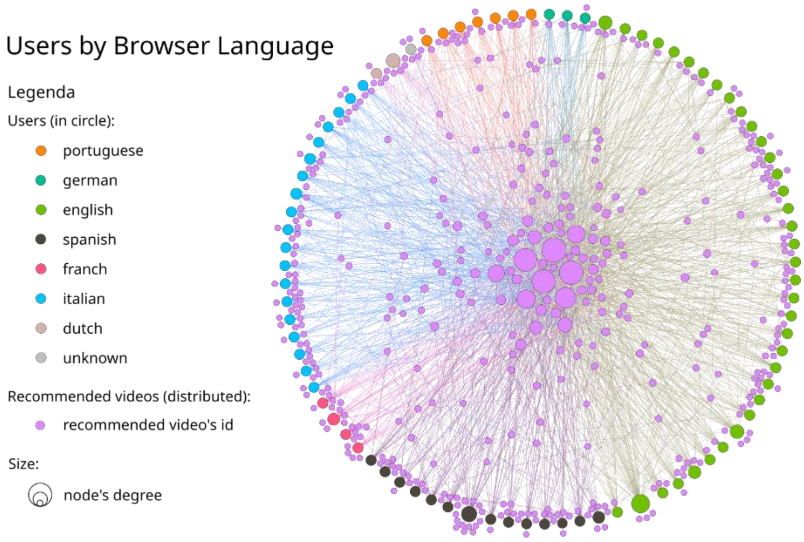


**Fig. 7.** Graph of the videos suggested to the participants while watching the video "How do I know if I have coronavirus? - BBC News"

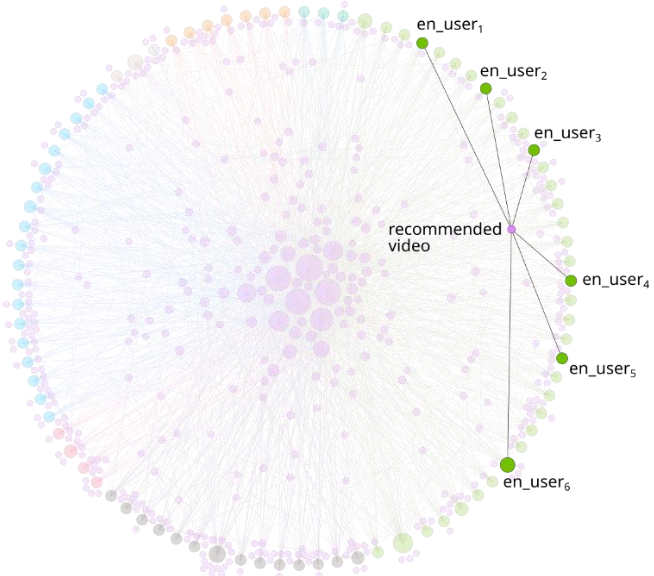


**Fig. 8.** Zoom of Fig. 7, an example of video suggested only to users with English interface.

In Fig. 8 we highlighted how some of the videos recommended appear only to users with English browsers. This shows that the participants in the experiment received personalized suggestions according to their characteristics, despite watching the same video. This type of analysis can demonstrate differences in the users' experiences tracing the most influential features that can generate changes in the platform experiences.

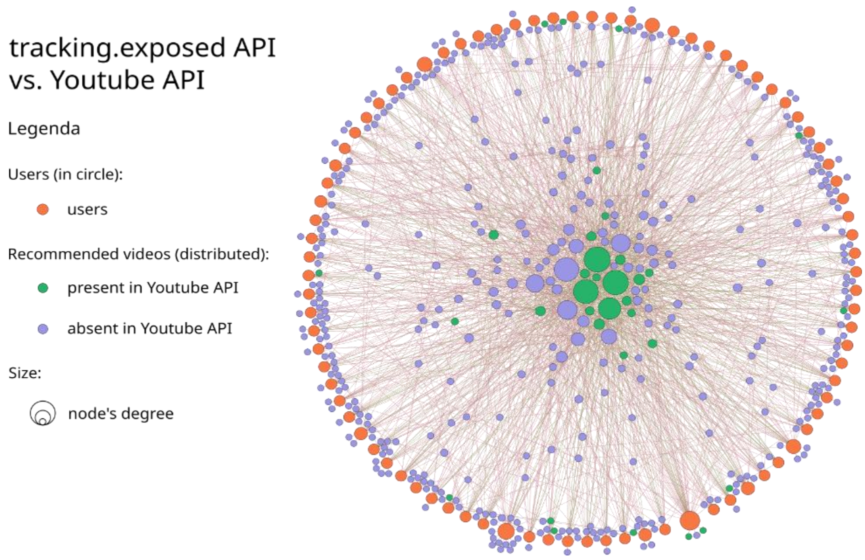


**Fig. 9.** Same graph of Fig. 7, here the colors highlight the differences between the videos recorded with Tracking Exposed and the ones retrieved with YouTube official API.

As we already said in the previous section, there is a huge difference between the recommended videos that we retrieved from the API and the actual recommendations (Fig.9). The majority of the videos retrieved by the Tracking Exposed tool are not present in the database created with YouTube's API. Some of the most suggested videos (biggest nodes in the center of the graph) neither. This is relevant because it is evidence against the usability of official YT's data in academic research. The official API cannot represent the real variance of suggestions present in the actual recommended videos. Many scientific articles [21, 22, 23] rely on this data, suggesting that those data can be used to explain the diffusion of videos on the platform, but according to our findings we might say that those data are just a generic representation of an ideal user that is really difficult to find in reality (no one of the users in our experiments gets the same recommendations as in the API).

The official API does not represent the various levels of personalization that occur in relation to the structural users' characteristics and to their past online behaviors. We cannot use API data to make inferences about personalization, polarization and filter bubbles, because those phenomena presuppose the study of real users in real context.

## 5 Conclusions

This research was intended as a proof of concept work, being the very first crowdsourced experiment carried out with our tool. It is not possible at this stage of the research to further generalize our findings on content analysis, as our sample was quite small (68 users). Nonetheless, we gained evidence of the existence algorithmic personalization, what is also known as filter bubble. Our analysis proves the necessity of further investigation on algorithmic personalization with crowdsourced and independent tools. In fact, our research also showed that official APIs cannot retrieve the majority of videos that are shown to the final users.

Further research on these themes should focus on 1) repeating the experiment with a larger sample 2) explore qualitatively the recommendations 3) study other structural characteristics of the users, understanding better the effects on the recommendations system, as we have already done with visualization time on similar recommender systems[9] 4) investigate the differences in the home page and in the search engine of the platform, as done recently by others groups[10] 5) compare the levels of curation in different languages, confronting the percent numbers of fake news or conspiracy videos.

## References


1. Fernandez, M., Harith A:, Online Misinformation: Challenges and Future Directions. In Companion Proceedings of the The Web Conference 2018, WWW '18, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 595–602 (2018).
2. Zollo F., Bessi A., Del Vicario M., et al.: Debunking in a world of tribes. PLoS ONE 12(7): e0181821. (2017).
3. Del Vicario M., Vivaldo G., Bessi A., et. al.: Echo chambers: Emotional contagion and group polarization on Facebook. Scientific reports,6:37825 (2016).
4. Pariser, E.: The filter bubble: What the internet is hiding from you, Penguin UK.P, (2011).
5. Zimmer, F., Scheibe K., Stock M., et. al.: Fake news in social media: Bad algorithms or biased users?. Journal of Information Science Theory and Practice 7(2): 40-53 (2019).
6. Bruns, A.: Filter bubble. Internet Policy Review, 8(4) (2019).
7. Covington, Paul, Jay Adams, and Emre Sargin "Deep Neural Networks for YouTube Recommendations". Proceedings of the 10th ACM conference on recommender systems. ACM (2016).
8. Zhe, Z., Lichan, H., Li, W., Jilin, et al. : Recommending what video to watch next: a multitask ranking system. In Proceedings of the 13th ACM Conference on Recommender Systems (RecSys '19). Association for Computing Machinery, New York, NY, USA, 43–51. (2019)


[9] Tracking Exposed, First coordinated observation of PornHub's algorithm: findings and how to let you reproduce the experiment. https://pornhub.tracking.exposed/potest/final-1/ (2019)

[10] https://www.their.tube/

9. Trielli D., Diakopoulos, N.: Partisan search behavior and Google results in the 2018 U.S. midterm elections. Information, Communication & Society DOI: 10.1080/1369118X.2020.1764605 (2020).
10. McKay, D., Makri, S., Guiterrez-Lopez, M, et al.: ACM, New York, NY, USA, We are the Change that we Seek: Information Interactions During a Change of Viewpoint. In Proceedings of ACM Conference on Human Information Interaction and Retrieval (CHIIR'20)., 10 pages. https://doi.org/10.1145/1234567890 (2019).
11. Robertson, R,, E., Jiang, S, Joseph, K., et. al.: Auditing Partisan Audience Bias within Google Search. Proeedings of ACM Human.-Computer. Interaction 2, CSCW, Article 148 (November 2018), 22 pages. DOI: 10.1145/3274417 (2018).
12. Hargreaves, E., Agosti C., Menasché D., et al.: Biases in the Facebook News Feed: a Case Study on the Italian Elections. International Conference on Advances in Social Networks Analysis and Mining, August 2018, Barcelona, (2018)
13. Arthurs, J., Drakopoulou, S., & Gandini, A. : Researching YouTube. Convergence, 24(1), 3–15 (2018).
14. Song, M., Yun, J., Anatoliy, G..: Examining sentiments and popularity of pro-and anti-vaccination videos on YouTube. In Proceedings of the 8th International Conference on Social Media & Society, pp. 1-8. (2017).
15. Abisheva, Adiya, David Garcia, and Frank Schweitzer. "When the filter bubble bursts: collective evaluation dynamics in online communities." In Proceedings of the 8th ACM Conference on Web Science, pp. 307-308.( 2016).
16. Bishop, S.: Anxiety, panic and self-optimization: Inequalities and the YouTube algorithm. Convergence, 24(1), 69–84 (2018).
17. Rieder, B., Matamoros-Fernández, A., & Coromina, Ò.: From ranking algorithms to 'ranking cultures': Investigating the modulation of visibility in YouTube search results. Convergence, 24(1), 50–68. (2018).
18. Sandvig C, Hamilton K., Karahalios K., Langbort C.: Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms. DATA AND DISCRIMINATION: CONVERTING CRITICAL CONCERNS INTO PRODUCTIVE INQUIRY, A PRECONFERENCE AT THE 64TH ANNUAL MEETING OF THE INTERNATIONAL COMMUNICATION ASSOCIATION. May 22, 2014; Seattle, WA, USA. (2014).
19. Bastian, M., Heymann, S., & Jacomy, M.: Gephi: an open source software for exploring and manipulating networks. In Third international AAAI conference on weblogs and social media. (2009).
20. Six, J. M., & Tollis, I. G.: A framework and algorithms for circular drawings of graphs. Journal of Discrete Algorithms, 4(1), 25-50. (2006).
21. Brbić, M., Rožić, E., & Žarko, I. P.: Recommendation of YouTube Videos. In 2012 Proceedings of the 35th International Convention MIPRO (pp. 1775-1779). IEEE. (2012).
22. Ledwich, M., & Zaitsev, A.: Algorithmic extremism: Examining YouTube's rabbit hole of radicalization. arXiv preprint arXiv:1912.11211. (2019)
23. Marchal, N., Au, H., & Howard, P. N.: Coronavirus news and information on YouTube. Health, 1(1), 0-3. (2020)