Algorithm exposed

Investigating Youtube personalization with yTREX

Team Members

Claudio Agosti, Davide Beraldo, Stefano Calzati, Zoé Charpentier, SeongIn Choi, Luca Del Fabbro, Anja Groten, Concetta La Mattina, Salvatore Romano, Giovanni Rossetti, Laura Swietlicki, Jeroen de Vos.

Contents

Summary of Key Findings

Briefly describe your most significant findings.

The experiment revolves around and identifies as the related content get personalized to different watchers, and, as a second goal, we check if there is a correspondence between what the official API returns and our experimental collection method.

1. We started by comparing the related suggestion for all the students in the class and they were different. We attribute the reason to the fact that they were using their everyday browser, and the evidence collected were personalized for their past activities.
2. All the class installed then a new browser (Brave) compatible with Chrome. With this we were sure to be able to compare all our activities without inherit any of our past online activities. When we did that, we could collect, visualize and reproduce graphically the difference between the perception of a profiled user and how YouTube treat someone anonymous (and with no past behavior).

During our research, we kept playing by comparing browsers in different conditions: logged in our personal YouTube account vs *clean-browser*, checking a video before and after a specific Google query, changing our charset, and watching the same video. The project focus can be better expressed as: "**We tried to control variables used by YouTube personalization algorithm in order to infer their weight in our individualized experiences**".

We found significant changes generated by the time watched per video and different strategies to suggest videos on the basis of the topic of the video watched. We compared the public information about a famous music video with our collection of data and the overlap is just partial.

The most interesting findings should drive further research on the topic with a bigger database. The first thing we are going to analyze in a greater depth is the qualitative and quantitative difference between the recommendations witch follow a political or an a-political content.

We will add some more features to the yTREX tool, to be able to record different test group in a specific folder, in order to make easier to control experiments with many hypothesis.

1. Introduction

YouTube provides us API to search for data on the platform, and a superficial explanation of the mechanisms that are used to personalize the user's experience. This resource is widely known by social media analyst. Our goal was to test a new approach.

The platform has been in the center of a long series of scandals in recent years. The increasing numbers of recommendations has been accused of driving users in to the "rabbit hole" of extreme contents[1]. Recently YouTube tried to give more apparent control on the personalization procedures by introducing the possibility to block some underiderable channels.

[1] https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html

Youtube took steps to *empower users with more control*[2], but this is not a systemic solution. Becca Lewis, an affiliate researcher at Data & Society and the author of a recent study about far-right content on YouTube, said to Wired[3] "**These changes seem positive at first glance, but they ultimately still put the burden of responsibility on the user, not on the platform**."

The digital giant also decided to reduce the visibility of borderline contents without taking the responsibility to take them off the site. And moreover, the automatic way of define which contents are dangerous, is really questionable and, at the end, it's not a public procedure, it's an algorithm black box (Lapowsky). We only have *some* information about the deep learning technique used to create the "Up next" list by the platform from a paper written by some of the main Google developers in 2016 (Covington, Adams and Sargin). On the other side, we have various research groups trying to shed light on the company, such as algotransparency.org, and we already have an idea of the different thematic groups actually present.

In general we can say that algorithms are governing and curating personal contents and news feed, on the contrary in traditional media content curation was done by human curators. One of the consequences that we can observe is that there is no restriction on the reinforcement of bad habits: if you like conspiracy videos, you will have more and more of them. That's happen because the system wasn't designed to give you something good, " it's built to get you addicted to" it (Maack).

In this project we used a browser extension developed as part of ALEX project. The tool is accessible to https://youtube.tracking.exposed, also referred as *ytTREX*. This approach allows us to record what YouTube offers to their consumers, and to study how the experiences become individualized by personalization algorithm.

Our final presentation is here available:
https://docs.google.com/presentation/d/1gUuLPZNB7LOofne_obbI716MGWF227Y64mCWbmiNQko/edit

2. Initial Data Sets

We built our dataset during the summer school, because our collection method works by watching at the same time the same YouTube video. We were **coordinating 10-12 people per time by opening the same video in the same moment.**
Using the API implemented in ytTREX , we were able to download all the participants observations. We based our research on these digitally assisted empirical observations.

We made one trial for each variable manipulated. That means that for every trial we have maximum 10 individual observation. Our dataset is not big enough to make inferences that we can generalize, but it was enough to compare how our observation makes our profiles diverge from an initial condition of relative anonymity, to a secondary condition where personal activities are artificially made to let Google study our profiles.

[2] https://youtube.googleblog.com/2019/06/giving-you-more-control-over-homepage.html
[3] https://www.wired.com/story/youtube-video-recommendations-changes/

3. Research Questions

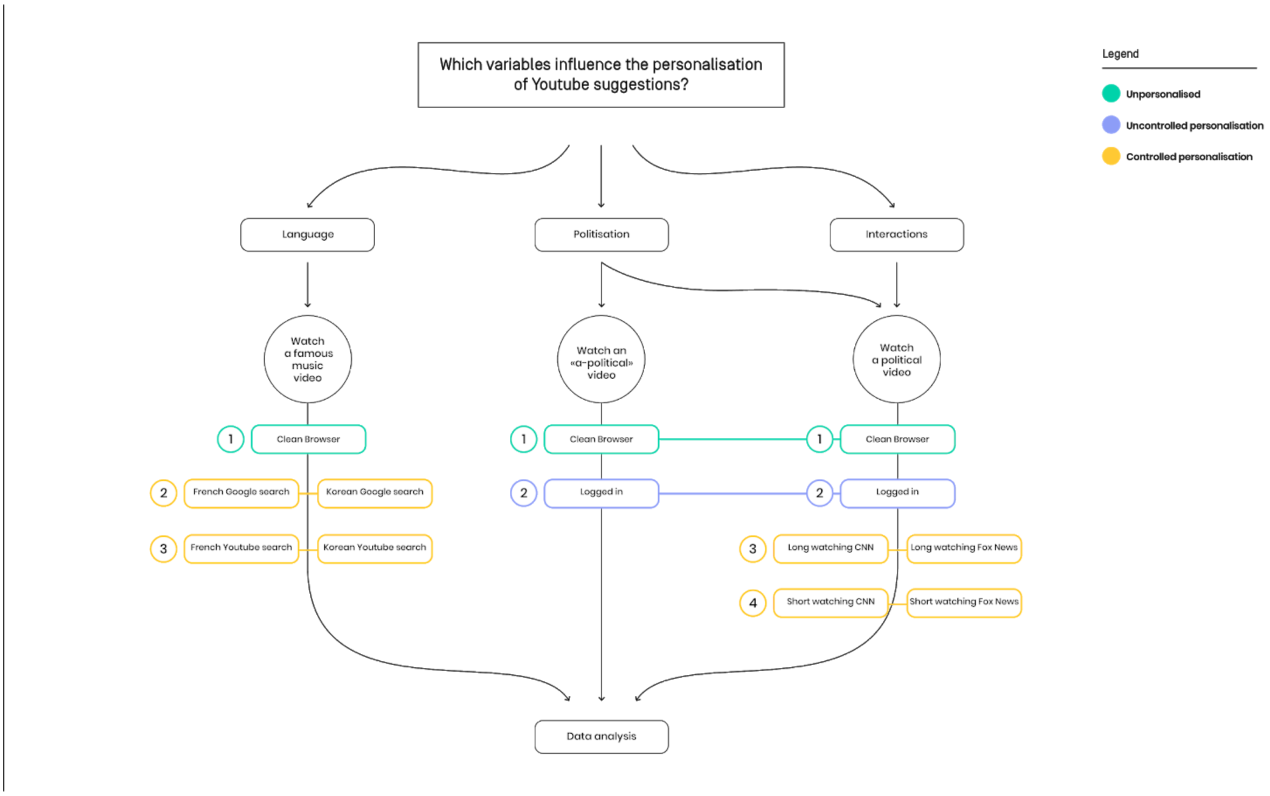
We tried to create a graphic representation of the personalization on YouTube, to be able to make some inferences also on the (journalistic definition of) “rabbit hole[4],” a more complex phenomenon.
We set up different trials to study, one by one, all the main variables involved in the personalization process. Our goal is to increase the awareness on the process that transforms user’s identity data and data extracted from the interaction with the platform, into personalized contents.
To do this, we didn’t rely on information provided by official API, but instead we used yTREX with “clean browsers” and then we compared the information obtained with the “officials”. We chose to make a front-end analysis, to be able to scrutiny results ourselves, rather than use the data provided by the company that we are studying.

On this basis, we formulated the **first questions related to the research context.** Are the API information really reliable and complete? But on the other hand: are the data obtained with single clean (o personalized) browser enough strong to be able to distinguish between personalization and randomness? Conscious that independent research on big platform is needed, are we able to define a really clean research set up?

In addition to those preliminary question, we were looking for different ways to represent graphically the personalization to be able to raise the awareness of the people regarding this issue.
With this spirit, we isolated different variables involved in the process, and we manipulated them in the minimum basic level, to show the big changes that follow the single act of giving a small information to the platform. **Which variable is playing a prominent role in the personalization process** between language settings, politicization of the content and user’s interaction with the platform? Which one is the most dangerous for its social and political implications?

4. Methodology

[4] https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html



We tried different setting to be able to compare the variation on related videos. Our aim was to test all the main variables implicated in the personalization of content. We made it into a carefully synchronized team effort, to try and test one variable at the time. The illustration above showcases our three principal levels of investigation:

1) **Language** through general setting and Google trackers.

2) Relevance of **politicization**: political vs a-political videos and Fox vs CNN.

3) **Interaction** with the platform: the time watched per video.

We started from the difference between a clean browser and our personal profile logged in. Then, we manipulated the information about the *language* changing the YouTube setting of “country” and “language”. In another trial we searched for 2 topics on Google and then we looked up the first three pages. We were divided in a French and a Korean query language groups. After this experimental manipulation, the two groups watched the same video to control if the algorithm uses the Google tracker information to create the “Up next” list .

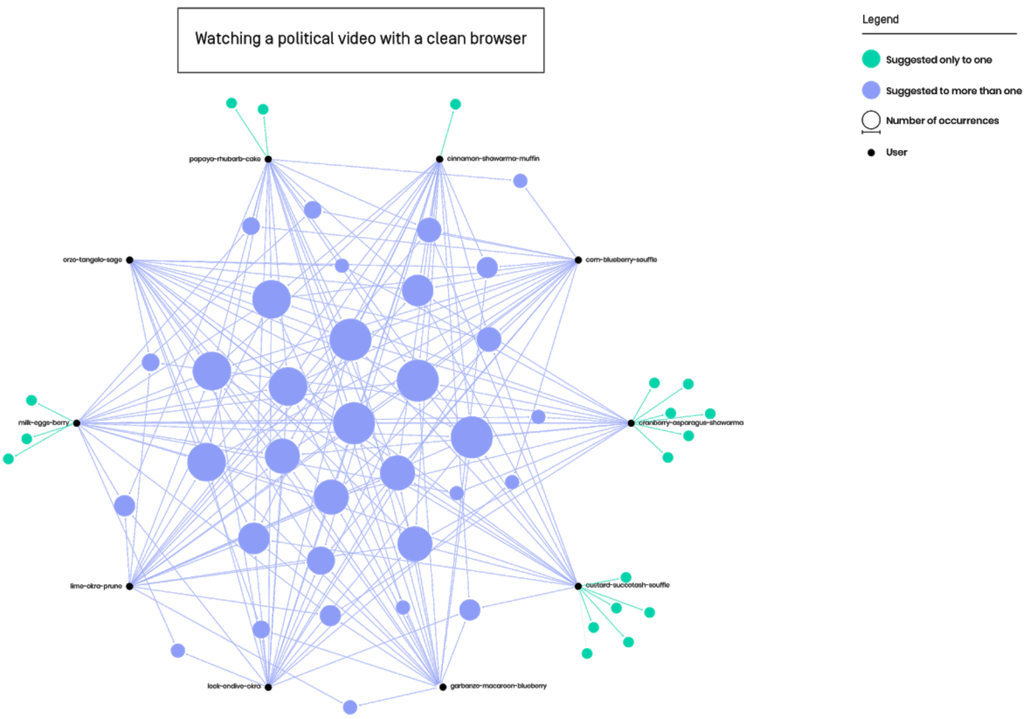
After the manipulation of the general settings, we moved to the type of content: we made one trial to see the differences between a *political* and an a-political video, watched with clean and logged browser. We also tested the diversity of related video network generated for two different users: the first one with just a Fox News video on its chronology, and the second with just a CNN video.

In the last part of the project we modify the *time of viewing* the same political video. To do this we replicated the experiment of Fox-CNN, but this time we diversified two test condition by watching for 20 second or 2 minutes the same clip. We don’t know how exactly Youtube recordes this data, for that reason we kept the interaction as basic as possible, avoiding any mouse movements. After 20 seconds or 2 minutes, we just closed the whole browser at the same time.

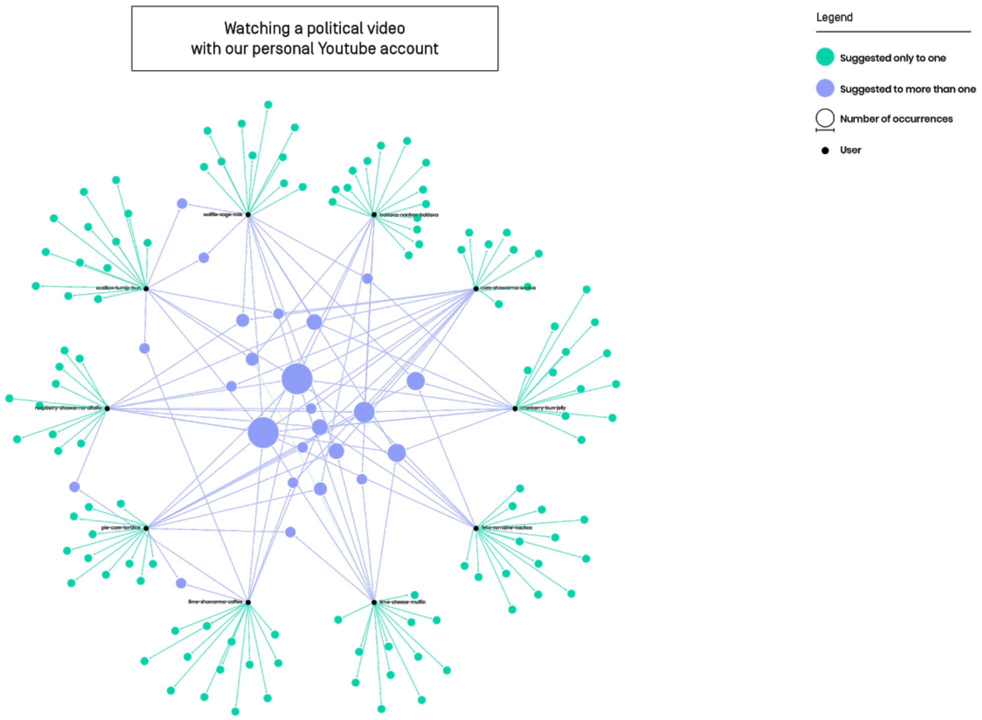
5. Findings

Describe your findings. Consider any counter-intuitive findings.

EXP 1: CLEAN vs PERSONAL browser. Can we graphically describe the YouTube personalization?

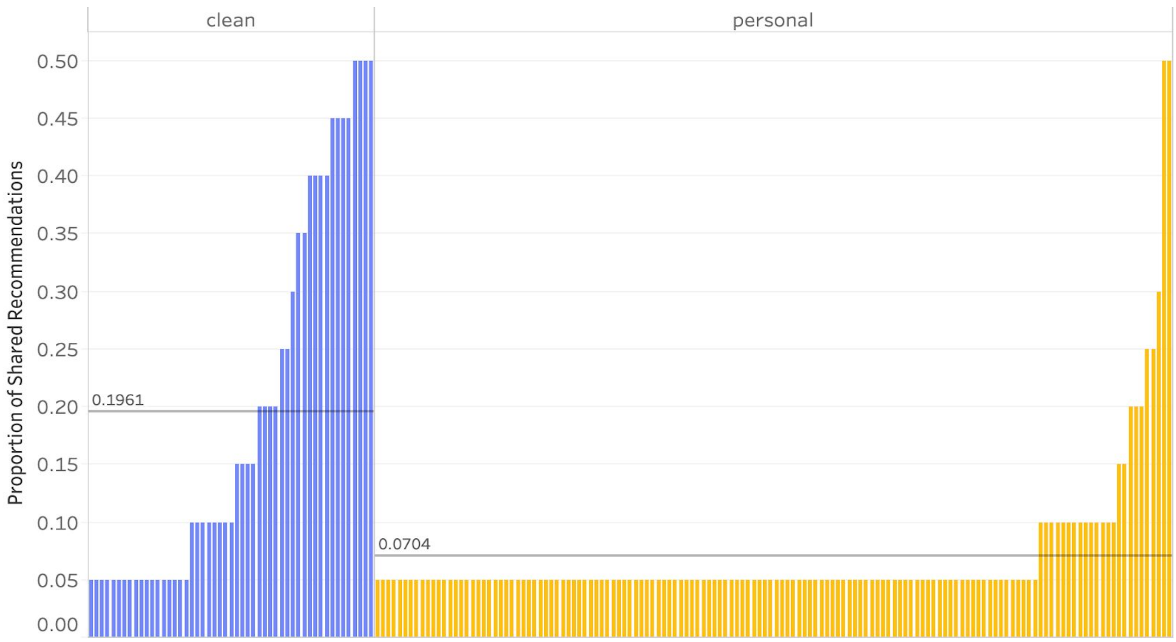


In the first image we can see 10 different users (black spots) using a clean browser without personal information. The related video (azur spots) show us that the users are sharing almost the totality of the suggested videos.



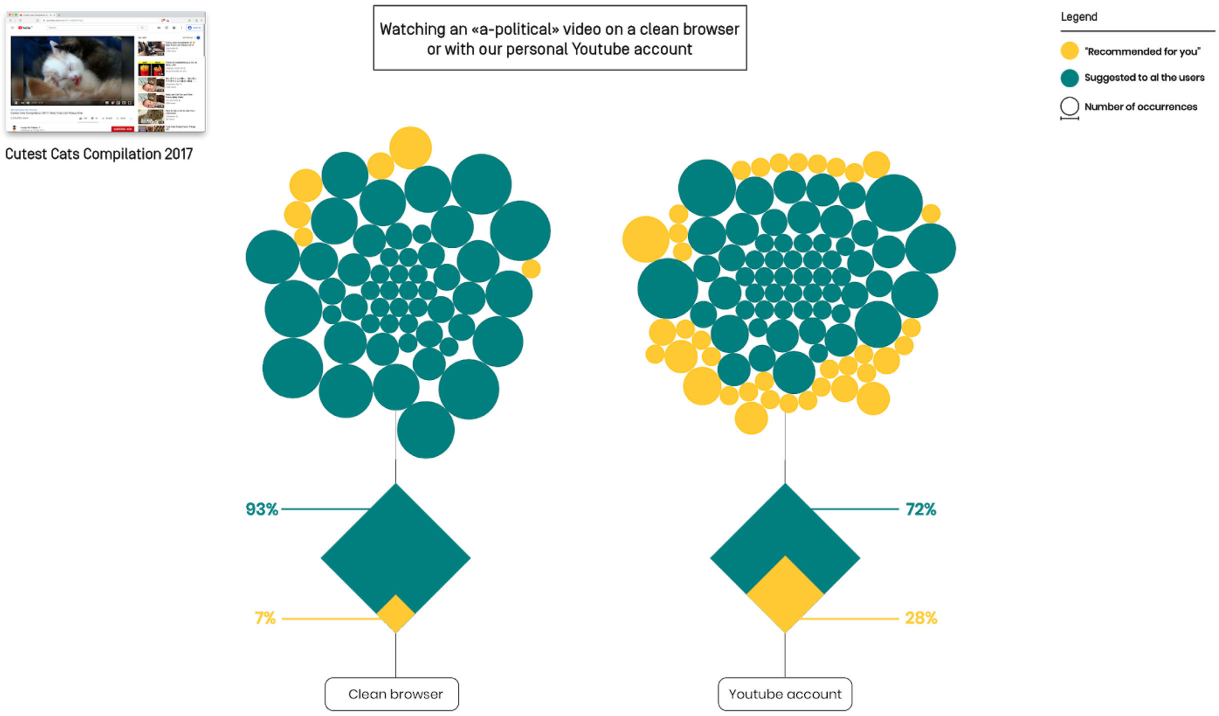
In the second image we can see the same video watched by the same numbers of users, but with personal account logged in, therefore with all our personal data on it.

In this trial the common suggested videos are less than in the previous trial. Note that the videos suggested just for a single user (and not for all the others) are much more.



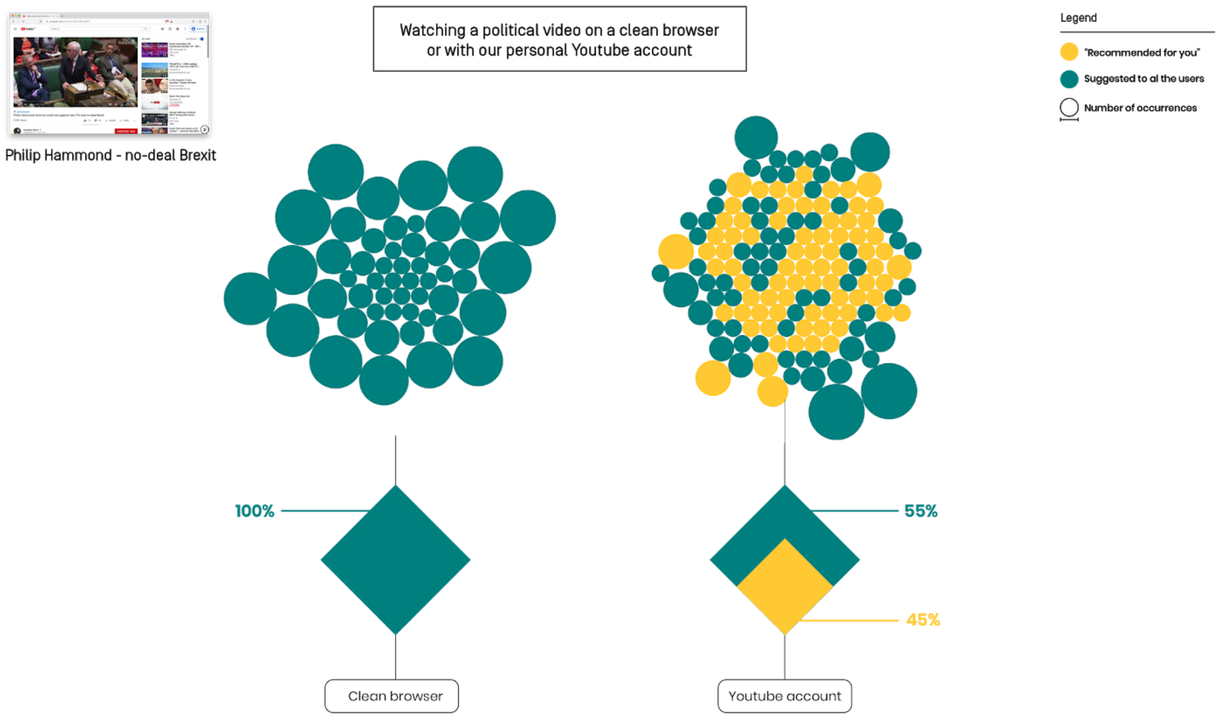
To quantify the difference, we first derived the proportion and mean of videos shared on a clean browser and a personal browser. When we visualize the proportions, we can see that the number of shared videos is larger in a clean browser compared to the number of shared videos on a personal browser. When a personal browser is used, the long-tail of single observationes is longer. Then, we conducted a **statistical test** to see whether the difference of mean and variance of proportion of shared recommendation is significant between the two groups. The results showed that mean and variance between the groups differed significantly at the p value of 0.001.

EXP 2: Recommended "for you" videos, POLITICAL vs A-POLITICAL. Does YouTube provide the same numbers of recommended "for you" contents on political and a-political issue?



We can see here the difference between the related videos (blue) and the recommended "for you" videos (yellow) watching an a-political content: "cutest cat compilation". In the first trial (left) we used a clean browser, in the second, we used our personal YouTube account.

The percentage of recommendations grows when the platform has more data on us. In this case we have some explicitly "for you" in the clean browser, but they are a minority.



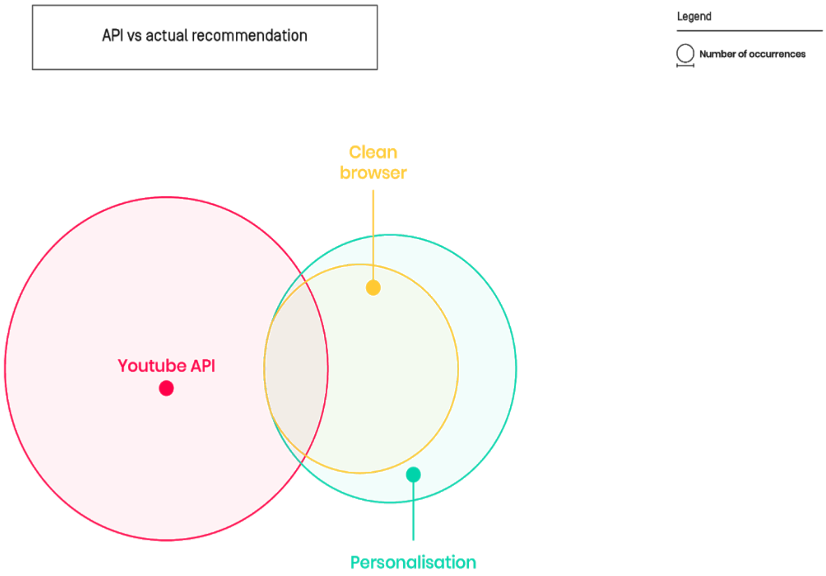
In this image we can see the same test of before, but with another video. We passed from an a-political to a political content about Brexit “Philip Hammond - do deal Brexit”. Here the differences between the clean and the personal browser set up are much more evident.

In this case we have no “for you” videos at all. It seems that Youtube doesn’t want to suggest anything on this sensitive issue, to be sure to do not make mistakes.

When we move to our personal accounts to see the same video, almost half of the contents are recommended “ for you”. The explanation could be that, when the platform has some data about the users, tries to personalized more the related video on political (an polarized) issue, than on the others.

To establish the reliability of this test we should have done a much bigger dataset. We can’t exclude that our personal profiles have more information about our political orientation. This might be why the numbers of the recommended videos increased. In the same way we can not be sure that, with different trails on different videos, we won’t have at least some suggestions “for you” also in the clean browser, as happened on the first trial.

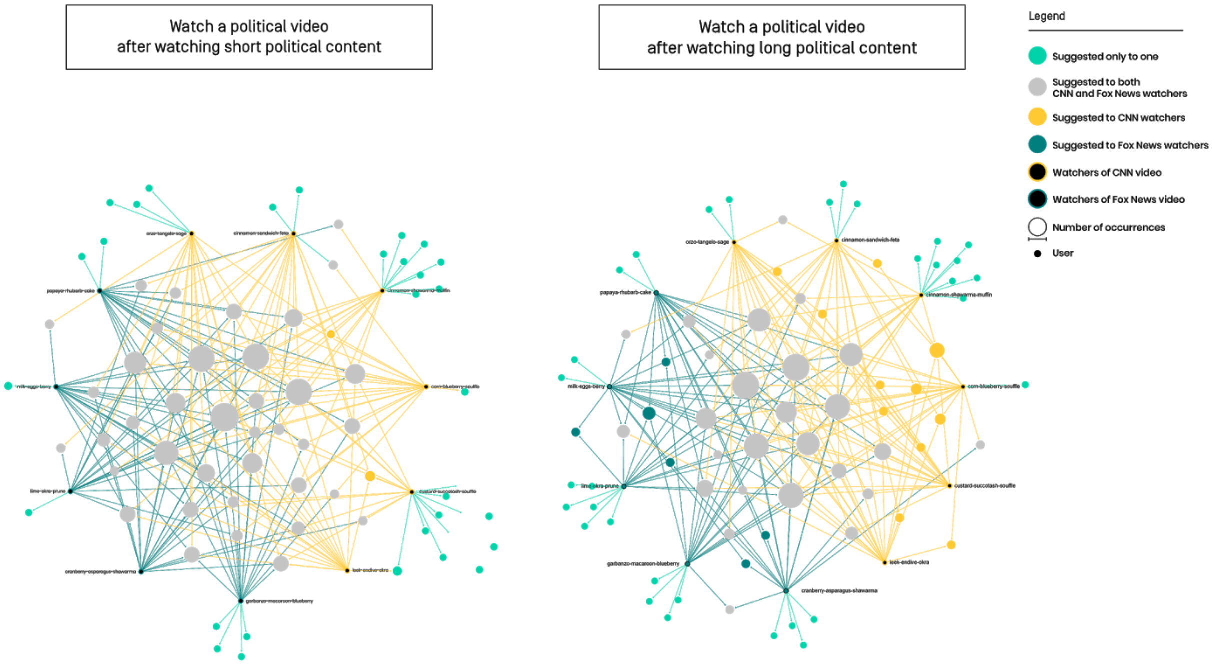
EXP 3: API vs ytTREX. Can we observe differences between the list of related videos provided by YouTube and the one obtained with our direct experience?



In this trial we tried to compare the related videos list provided by API on a really famous video ('Gangnam Style') with the related videos that we can observe using ytTREX, differentiating personal and clean browser.

It seems that the list published by the platform overlaps with just a subset of common videos with list that we can see directly. Can we really trust on data provides by platforms? With a larger set of videos and with larger numbers of related video collected, we can try to answer this question, using ytTREX.

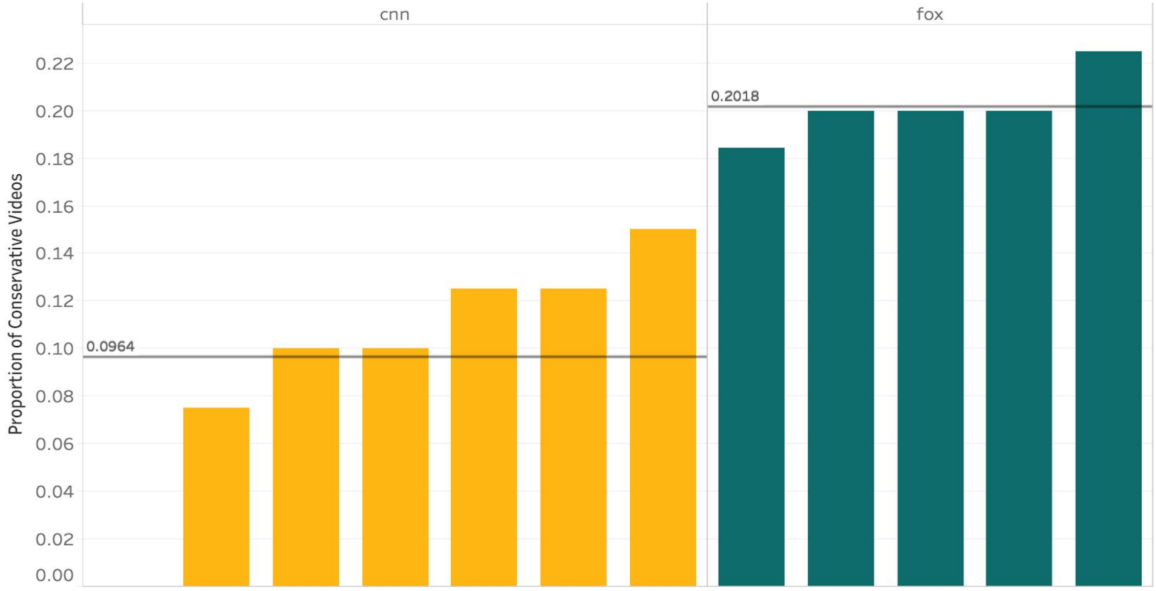
EXP 4: 20 seconds vs 2 minutes watch. Testing the influence of 'interactions' on personalisation.



We observed the different levels of personalization generated by watching the same video for 20 seconds or 2 minutes. In this trial we divided ourselves into two different groups, the first one has seen a Fox News video and the other have seen one clip from the CNN channel. After this manipulation, we have seen the same political video about Brexit and we compared the related videos results.

As we can see in the graphs, watching Fox and CNN for 20 second create just three common suggestions specific for CNN watchers (yellow spots) and no one for the other group (blue spots). In fact we can see a lot of videos suggested to both groups (grey spots) and some suggestions just or the single user (azur spots).

In the second trial, extending the watching time till 2 minutes, we can see eleven common suggestion to at least two CNN users, and six for Fox.



The difference between the two groups of related contents is significantly different, and we tested it using the ANOVA in R. Before running the test, we re-coded the sources according to their political views. Then we calculated the proportion of conservative videos shared, and the mean and variance. As seen in the bar graph, the mean of proportion of conservative videos shared in the Fox group is higher than the CNN group. **The ANOVA test** revealed that the difference in mean and variance between the two groups was significant at the p value of 0.001. That means that the algorithm suggests more conservative videos to the Fox users. The suggestions don't seem to be related only to the argument of the clip, but also to the political orientation of the channel.

6. Discussion

Our aim was to test the ytTREX tool because we were the first group to use it for research proposal. That's why we wanted to have a trial for as many variables as possible to be able to create the "clean browser" methodology, indispensable for further research.

Despite everything, we have shown how different variables affect the personalization of contents. Graphic representation of the differences created by a **really small modification of the information related to our "clean browser", creates quite big differences in the related videos.**

Some authors said that YouTube usually increases the visibility of inflammatory contents, and they focus their attention on the problem of the propagation of these kind of content also to the mainstream watchers (Rieder, Matamoros-Fernàndez, and Coromina).This research has been an attempt to provide empirical data to make claims about the personalisation mechanism present on Youtube. Opposed to more popular claims regarding the inherent polarisation of personalisation, this research and yTREX allow to start exploring the claims made.

We need to investigate the real effect of the algorithm across different topics, to see what kind of controversy verity increases the type and number of times it is suggested. We have seen significative differences across different subjects in particular between political and a-political videos, but we need to go deeper in the analysis to have a more complete view.

To do this we need to be careful in the use of Yt API because it might not be reliable (Exp3), even if harvesting could be more timedispending, is the best way to make independent research, and to not be dependent on the platform studied.

Contrary to expectations, our minimal manipulation of the language settings through the visualization of 3 different pages on Google did not give us visible results. In further research we will need to apply a more heavy handling to see how many Google tracker's inputs Youtube needs to change the list of related videos. Changing the genearles Youtube settings we noticed some differences, but not really relevant. Some of the videos had translated titles, but the videos suggestied were similar. Probably the company uses this general feature just to make the interface more accessible, and not to personalize the contents.

Even if the **language** does not seem to be relevant for the type of the videos suggested, the **interactions** with the platform are surely used to personalize user's experiences, but we need to deeply understand how. The **political perspective** of a video watched is really important to generate the suggestions "for you", and this is the variable with more social implications to deepen.

7. Conclusion

We have many challenges in front of us and many questions to answer. Now that **we have a clean browser methodology**, and we tested the tool in a research context, we need to focus on some analysis based on a bigger database and a deeper and more specific issue.

Surely, the most interesting difference is between political and a-political videos. The differences on personalization can be not relevant if the topic is "funny cat" for example. On the contrary, when we talk about politics, conspiracy, fake news, and similar, this apparently neutral algorithm becomes part of the political life of our society.

We already know, by the admission of the developers [Convington], that the time watched per play is one of the most relevant factors used to create the suggested list. Actually, it is one of the best ways to understand the appreciation for a video and than to figure out if a suggestion has been effective or not. It's important to discover how exactly YouTube records this data to be able to consider this variable in future studies. **It's possible that there are other types of interactions recorded by the plattform** like mouse dynamics, but we need further research to ascertain it with certainty.

8. References

Algorithms Exposed, https://algorithms.exposed.

Algotransparency, https://algotransparency.org.

Covington, Paul, Jay Adams, and Emre Sargin “Deep Neural Networks for YouTube Recommendations”. *Proceedings of the 10th ACM conference on recommender systems*. ACM, 2016, https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/45530.pdf.

Lapowsky, Issie “YouTube Will Crack Down on Toxic Videos, But It Won't Be Easy.” *Wired*, 25 January 2019, https://www.wired.com/story/youtube-recommendations-crackdown-borderline-content/.

Maack, Már Másson “‘YouTube recommendations are toxic,’ says dev who worked on the algorithm.” *The Next Web*, June 2019, https://thenextweb.com/google/2019/06/14/youtube-recommendations-toxic-algorithm-google-ai/.

Matsakis, Louise “YouTube Is Giving You More Control Over Video Recommendations ”*Wired*, 26 June 2019, https://www.wired.com/story/youtube-video-recommendations-changes/.

Rieder, Bernhard, Ariadna Matamoros-Fernández, and Òscar Coromina “From Ranking Algorithms to ‘Ranking Cultures’: Investigating the Modulation of Visibility in YouTube Search Results”. *Convergence*, vol. 24, no. 1, Feb. 2018, pp. 50–68, doi:10.1177/1354856517736982.

Other articles:

Rose “The Making of a YouTube Radical.” *The New York Times*, 8 June 2019, https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html.

Tufekci, Zeynep “Algorithms Won’t Fix What’s Wrong With YouTube.” *The New York Times*, 14 June 2019, https://www.nytimes.com/2019/06/14/opinion/youtube-algorithm.html.

Tufekci, Zeynep “YouTube, the Great Radicalizer.” *The New York Times*, 10 March 2018, https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html.