Characterizing videos, audience and advertising in Youtube channels for kids

Camila Souza Araújo*[1], Gabriel Magno*[1] Wagner Meira Jr.[1], Virgilio Almeida[1,2], Pedro Hartung[3], and Danilo Doneda[4]

[1] Universidade Federal de Minas Gerais, Belo Horizonte, Brazil,
{camilaaraujo, magno, meira, virgilio}@dcc.ufmg.br
[2] Berkman Klein Center, Harvard University, Cambridge, USA
[3] Harvard Law School, Cambridge, USA,
[4] Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil,
danilo@doneda.net

Abstract. Online video services, messaging systems, games and social media services are tremendously popular among young people and children in many countries. Most of the digital services offered on the internet are advertising funded, which makes advertising ubiquitous in children's everyday life. To understand the impact of advertising-based digital services on children, we study the collective behavior of users of YouTube for kids channels and present the demographics of a large number of users. We collected data from 12,848 videos from 17 channels in US and UK and 24 channels in Brazil. The channels in English have been viewed more than 37 billion times. We also collected more than 14 million comments made by users. Based on a combination of text-analysis and face recognition tools, we show the presence of racial and gender biases in our large sample of users. We also identify children actively using YouTube, although the minimum age for using the service is 13 years in most countries. We provide comparisons of user behavior among the three countries, which represent large user populations in the global North and the global South.

1 Introduction

All over the world, digital technologies are shaping children's lives for better or for worse. Policy makers, researchers and educators who work with children's rights agenda recognize the social impact of digitization for young people's lives [4, 8, 9, 5, 13]. Online video services (e.g., YouTube, Netflix, BBC, etc.), messaging systems (e.g., Whatsapp, Messenger, etc.), games (e.g., Apple, Google Play, IGN, Gamespot, etc.) and social media services (e.g., Snapchat, Facebook, etc.) are tremendously popular among young people in many countries [11]. YouTube in particular has been viewed as an alternative to traditional children's TV[6]. Millions of children are already watching videos on YouTube, most of them logged in from their parents' accounts. For example, the channel of a popular youtuber, Joseph Garrett, has 7.8 million subscribers and its videos have

* These authors contributed equally to this work

been viewed 5.3 billion times, making it one of the most popular British YouTube channels for children [7]. Most of the digital services offered in the Web are funded by advertising, which makes advertising ubiquitous in children's everyday life. YouTube offers ad-funded video channels. As a consequence, several questions about the role of advertising in children's life arise. What forms of advertising do children face on the internet? How do children react to online advertising? The goal of this paper is to provide a detailed quantitative characterization of users, videos and advertising in a sample based on several popular YouTube channels for kids in US, UK and Brazil.

There are several types of ads. For example, advergames are video games created by a company with the intention of promoting the company itself or its products. Usually these games are distributed freely as a marketing tool. There are cases of food and drink companies that target children with advertising unhealthy products on various internet platforms. An European Commission study [11] reports that online marketing to children and young people is widespread, and in some cases various marketing techniques used are not always transparent to the child consumer. There are marketing strategies that target children on YouTube with advertising disguised as other content. They use popular youtubers to pitch products and brands as non-commercial content in videos that are viewed worldwide. Characterizing and understanding these strategies and their effectiveness is a key task for making the internet and the web a better place for children.

Recent figures published by ITU (International Telecommunication Union) in 2016 show that developing countries now account for the vast majority of internet users, with 2.5 billion users compared to one billion in developed countries. According to [14], one of every three internet users in the world is a child. internet is becoming the main medium through which children collaborate, share, learn and play. In order to understand rights, risks and opportunities for children on the internet, it is important to look at countries from both the global North and the global South [9]. Because of its worldwide penetration, YouTube channels for kids is a good scenario to understand advertising campaigns that target children. The paper provides a study of interaction of online advertising and Youtube for Kids audience in US, UK and Brazil. The survey on digital marketing by the company GroupM[5] in April 2017, estimates 44 million YouTube users in UK and 72 million in Brazil. The Statistics Portal[6] estimates 180 million YouTube users in US.

In order to collect and analyze YouTube usage data, we developed an experimental methodology based on the combination of free APIs and open source tools available on the internet. The results of the characterization presented in this paper can be useful for policy makers in different countries to assess the need of public policies to protect children online. To the best of our knowledge, this paper is the first one to study and characterize videos, audience and advertising in internet channels for children.

[5] www.groupm.com
[6] www.statista.com

Overall, we make the following contributions:

1. We develop a simple experimental methodology to collect and analyze large amounts of YouTube usage data based on APIs and open source tools available on the internet.
2. We integrate free text-analysis and face recognition tools to identify age, race and gender of YouTube channel users as well as to characterize the behavior of those users.
3. We identify children actively using YouTube, although the minimum age for using the platform is 13 years, according to their Terms of Service. Even if some usage of under 13 is generally considered as fair due to parents' or legal responsible consent and supervision, the fact is that if children are actually using the platform they can be exposed to advertising, what raise concerns about compliance with publicity regulation in several countries.
4. We show the presence of racial and gender bias in the large sample of YouTube users in our data sets. The percentage of black users is very small when compared to white and Asian users.
5. We analyze the behavior of YouTube users in three countries, US and UK in the global North and Brazil in the global South. We show differences and similarities in the demographics of YouTube channel audience as well as in the categories of products and brands associated with the videos of the channels.

The rest of the paper is organized as follows. We begin with a description of research questions associated with online advertising for children in section 2. In section 4, we discuss the computational approach used to gather and analyze data from different internet channels for kids. A detailed description of the datasets collected from YouTube channels is given in section 5.1. Next, in sections 5.2 and 5.3, we characterize videos and advertising of popular YouTube channels. Finally we describe and characterize the behaviors of users of YouTube channels in section 5.4. Section 6 summarizes our findings and discusses future work.

2 Research Questions

In this section we discuss the research questions that we address in our work. The expansion of the use of social and digital media led to the expansion of the presence of marketing to children through digital platforms. As mentioned, advergames, product placement in YouTube videos and online games, marketing in social networks and other strategies are commonly used by companies to attract the attention of children and persuade them to consume certain products or services. However, unlike traditional media, marketing in the digital environment takes new forms and many of them are more difficult to be clearly identified.

By providing a detailed characterization of YouTube channels for kids, this paper aims at shedding some light on streaming video-on-demand programming that target children all over the world. It also seeks to understand how the children's audience interacts with channels and videos through the children's engagement in the conversations in the video comments in YouTube[4]. As a consequence of our research goal, we ask the following questions:

- What are the characteristics of the content of the most recurrent videos on children's channels?
- What does characterize the audience to the videos for children (e.g., is there a predominance of gender in the audience and also in the young youtubers)?
- Which classes of products are marketed to a specific target (i.e., gender, age, ethnicity)?
- Is it possible to measure the percentage of children's audience in the YouTube channels examined?
- What are the gender, specific age and ethnicity among the children identified?
- What is the content of the most recurrent videos on children's channels?
- Is it possible to identify publicity directly aimed at children on the channels?

In order to investigate the stated research questions, this study relies on data collected from popular YouTube channels in US, UK and Brazil. We developed a computational methodology based on open source code to analyze the data.

3 Related Work

We now briefly summarize existing studies related to YouTube analysis, as well as studies of user behavior in Social Media.

Online social networks are popular platforms for people to connect and interact with each other [15]. According to Benevenuto et al. [3], understanding users behavior on social networking sites creates several opportunities. For example, accurate models of user behavior are important in social studies and viral marketing, since viral marketers may exploit models of user interaction to spread their content quickly and widely. Nowadays, many children use the Internet and mobile technologies as part of their everyday lives. The overlap of the online and offline world comes with a range of digitally-mediated opportunities and risks. Reference [13] provides a qualitative analysis of different social media sites to assess if they provide healthy environments for children and teenagers. In [17], the authors investigate the effectiveness of Internet filtering tools designed to shield teenagers from aversive online experiences. Based on $1,030$ in-home interviews conducted with early teenagers aged from 12 to 15 years, the paper shows that Internet filters were not effective at shielding early teenagers from aversive online experiences, that include scary online videos. Magno et al. [12] studied the Google+ environment and compared its network structure with Facebook and Twitter, and noticed that it has a higher average path length and higher reciprocity. They also compare the user profile characteristics between different countries, and found that some countries are more private than others.

As pointed out by Benevenuto et al. [2], online video sharing systems have been increasing and gaining popularity. These environments allow several kinds of interactions between users and videos, such as publication of comments. This is the most related work to ours in terms of characterization methodologies, however the authors presented an in-depth workload characterization of sessions and requests on an video server different than YouTube. In terms of sociological studies of the impact of YouTube on kids, according to the authors of [18],

the evidence on how use of the Internet impacts on child rights and well-being is still scattered and patchy in most countries. Livingstone and Local [10] discussed the problem of audience measurement techniques regarding children's television viewing, because of the diversification in devices on which television content can be viewed. It is already well known that television content can be viewed on Internet-enabled devices and Internet content can be accessed via Internet-enabled television sets, but such viewing cannot be measured satisfactorily at present.

There are few articles in the literature that address quantitative analysis of online advertising for children in Internet video channels. For example, Dehghani et al. [5] uses data collected from Italian students to analyze the perception of YouTube by young people. Data were obtained from 315 questionnaires. The results show positive aspects of YouTube in terms of entertainment, informativeness and customization. The negative aspect is related to YouTube advertising. Unlike this reference, our paper relies on a large scale datasets collected from YouTube channels.

4 Experimental Methodology

In this section we describe the computational approach adopted to answer the aforementioned research questions.

4.1 Rationale

The research questions outlined in Section 2 are hard to verify and and quantify, mainly considering that most of our input data that are publicly available and are composed of free text and images. One immediate consequence is that it is unfeasible, in practice, to get fully accurate and complete datasets about the video body we want to analyze. Then, we adopt an approach based on identifying evidences. Each evidence makes explicit a piece of information about the entity being analyzed. Considering YouTube users, examples of evidences are his or her gender, age and race, as extracted from a profile picture. In this work, we chose a set of evidences that demonstrate the occurrence of advertising in child-oriented videos, as described in Section 4.2.

These evidences should be conservative, although it is possible to improve the gathering techniques and be able to identify evidences. For instance, when we label a video as an advertising piece, we should be as sure as possible that it really is. The immediate impact of our approach is that all our figures are lower bounds of the actual evidence counts. Although we are usually not able to perform analyses that demand accurate counts, they demonstrate clearly the occurrence of targeted phenomenon or behavior. It is important to emphasize that our strategy also leverages on the fact that there are already a huge number of techniques and tools that may be promptly used for identifying evidence, as we discuss next.

4.2 Methodology

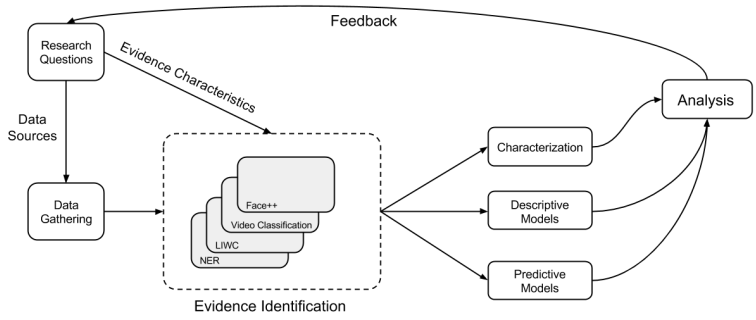

Fig. 1. Methodology

In this section we present our methodology for assessing the occurrence of not only advertising in YouTube videos, but may also serve to analyze the occurrence of various phenomena associated with Internet applications. As we discussed in Section 2, we look for identifying and modeling three groups of evidences: (i) content characterization; (ii) audience profiles; and (iii) detection of products and their publicity in videos.

The starting point of our methodology is the set of research questions we want to answer. We then select the data sources from which we will gather data for each question. We also map the questions into evidences to be identified, which demand the application of one or more techniques to the data, usually enhancing them. Examples of evidence employed in this paper are the positivity of video meta-data and user profile inferred based on face snapshots, which add attributes to both video and users, respectively. The evidences may be characterization findings, descriptive models or predictive models. Characterization may use summary statistics, among other techniques, to detect invariants, trends and other properties. Descriptive models comprise patterns and models inherent to the data, such as clusters and correlations. Predictive models estimate samples' class or numerical dependent variables. It is worth mentioning that there is a large spectrum of techniques for identifying and modeling evidences, most of them freely available in the Internet. The enriched data, characterization findings and derived models are then used for analysis and answering of the original research questions. The last step is to, considering the correctness and completeness of the answers, improve the whole process towards increasing its quality. Figure 1 depicts the methodology proposed.

In the next sections we discuss the techniques used for evidence identification and modeling in more detail, as well as how they help our analysis.

4.3 Data Gathering

YouTube is a large-scale video sharing online platform where users can produce and/or consume content. On YouTube users need first to create a channel to upload videos. Users do not need to be logged in to watch videos, but they need

to be logged in to comment and 'like' videos. We collect information of YouTube videos and comments using Google's YouTube Data API [7], accessing it directly with Python 3 scripts. The data collection process was performed in 5 phases:

1. **Select list of channels**: we manually select a list of 41 popular YouTube channels targeted to children. This selection was performed by children's rights experts and based on their popularity and also empirical evidence they may employ advertising strategies.
2. **Retrieve list of videos**: for each channel, we collect its list of videos. Due to API limits, we only get the last 500 videos published in the channel.
3. **Retrieve video statistics**: for each video, we collect its information and statistics.
4. **Retrieve comments**: for each video, we gather the comments published by users about it.
5. **Retrieve replies**: YouTube users may reply to a video comment, thus for each comment we collect its list of replies as well. In our analysis we handle replies as normal comments.

As we discuss later, the collected information allows extensive analysis about the video characteristics, the marketing strategies it may employ, and the observed impact of such strategies.

4.4 Evidence Identification and Modeling

In this section we present the various strategies and techniques used in this study for sake of evidence identification and modeling.

Entity Recognition In order to characterize the videos we need to extract entities (names, brands, products etc.) that are mentioned in it. We use a technique called Named Entity Recognition (NER), a method that labels sequences of words into categories of things, such as company names, person and cities. We use the Stanford NER [8] tool with the English pre-trained model. Unfortunately, it does not have a pre-trained model for Portuguese, so we use this technique only for the videos of U.S channels. For both Portuguese and English, we also detected entities, in particular products and brands, by assessing the video meta-data, as discussed in Section 4.4.

Sentiment Analysis We assess the public perception on the videos published by analyzing the content of the comments written about them. To attain this task, we use the Language Inquiry and Word Count (LIWC) [16], a lexicon used to verify the occurrence of words from several grammatical (e.g., pronouns, verbs, and articles, among others) and semantic (e.g., positive emotion, social, motion) categories. We use the LIWC 2007 dictionary, whose complete list of categories and word examples is available at LIWC's website [9]. For sake of our analysis in this work, we calculate the proportion of comments that contained at least one word of a particular category and label the comment to all categories that match.

[7] `developers.google.com/youtube/v3/`

[8] `nlp.stanford.edu/software/CRF-NER.shtml`

[9] `liwc.wpengine.com/`

Product Category Identification A key component for our analysis is the identification of products and respective products categories that may be marketed and advertised in the videos being analyzed. However, watching and labeling thousands or millions of videos is unfeasible. We adopt the strategy described next for identifying product categories present in each video. We start by extracting the tags of the "video tags" field for each video, which is a list of labels manually inserted by the owner of the channel. For the videos of U.S channels, we also consider the entities mentioned in the "description" field, which are extracted using the NER tool (as described in Section 4.4).

The next step is, given the list of tags (manual tags and NER tags) for each video, we calculate the frequency of the tags among all videos, and compile a sorted list of the most popular tags for each country. Then, we retrieve the top 1,000 tags in each list and manually check the assignment of each tag to one of the 23 categories of products presented in Table 1. If the tag does not match any of the categories, we ignore it.

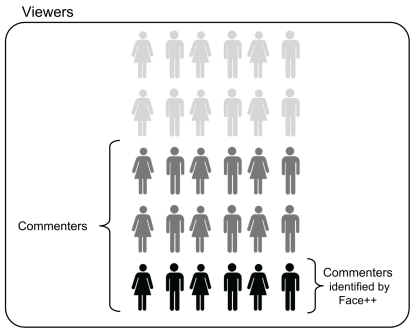

Fig. 2. 'Viewers' are all users who watched the video and 'commenters' are users who left comments. We still have a subset of commenters for whom Face++ has identified a face.

Product Categories		
Footwear	Water	School Suplies
Clothes	Snacks	Electronics
Fast Food	Fresh Food	Pet Products
Chocolate	Food (other)	Books
Candies	Cosmetics	Travel and Recreation
Sodas	Make Up	Services
Juices	Toys	Movies and Shows
Yogurts	Games	

Table 1. Categories of Products Used to Classify Videos

Video Classification Once we assigned tags to product categories, we may classify the videos by simply verifying whether it contains one of the tags of a particular category. It is important to notice that a tag might have been classified into more than one category (e.g "Disney toys" is both from "Toys" and "Movies and Shows" category). In the same sense, a video might be assigned to two or more categories. Notice that the process we employed provides good precision, but not necessarily a good recall, since it relies on the channel owner, who provides the tag definition and description information. It is beyond the scope of this work to assess how complete, accurate, and consistent across videos and channels these data are.

User Visual Profiling The user visual profiling aims to determine user information such as gender, age and race of the users who comment the videos. As

previously stated, after selecting a channel list, we collect information from the last 500 videos of each channel, which includes the URL of the YouTube profile image of all users who left comments. We then download the profile pictures associated with all users and use Face++[10] to extract information such as age, race and gender about each face in the photo. Face++ is an online API for facial recognition and its accuracy is known to be over 90% for face detection [1]. It is important to note that not all users use real photos as a profile image, so Face++ is not always able to identify a face. In the next section, we will present the number of identified faces that composes our dataset. It is important to mention that, although we employed just visual profiling, any technique that provides such information may be used. The key issue here is the coverage of the profiling information acquired considering the user population and their accuracy.

5 Data Analysis and Results

5.1 Datasets

Now we present a brief characterization of the datasets collected for this work. We chose to collect data from 24 Brazilian YouTube channels, and 17 YouTube channels for kids produced in English from United States and United Kingdom. The rationale for selecting US, UK and Brazil is the following. Most of the YouTube content is produced in English[11]. However, Brazil is the second largest market considering time spent on YouTube[12]. The three countries represent a large number of YouTube users in the global North and global South. Our Brazilian dataset comprises data about 7,664 videos and 10,940,565 comments associated with them, issued by 2,982,595 distinct users. That is, the same user may leave more than one comment. It is important to emphasize that throughout the work when we refer to 'users' we do not refer to all users who watched the video, we refer to the subset of users who made comments, as shown in the Figure 2. From now on, we will call those users - users who left comments - of 'commenters', in this way we avoid confusing them with the users who only watch the videos. Of this total of commenters we were able to extract information of 129,286 faces. Table 2 summarizes the size of the dataset collected from the channels of both countries. In Table 3 (Appendix) we summarize the dataset composition.

	#channels	*#videos*	*#views*	*#comments*	*#commenters*	*#faces*
Brazil	24	7,664	4,614,161,928	10,940,565	2,982,595	129,286
US+UK	17	5,184	37,401,690,211	3,569,553	2,013,419	9,248

Table 2. Dataset Summary

[10] www.faceplusplus.com/

[11] medium.com/@synopsi/what-youtube-looks-like-in-a-day-infographic-d23f8156e599

[12] googlediscovery.com/2017/03/23/google-brasil-dados-importantes-sobre-o-google-no-brasil/

5.2 Analysis I: Videos and Channels

In this section we will present a characterization of the videos present in our dataset. Figure 3 shows the number of comments and commenters - users who have left comments - per channel. In general, Brazilian videos have more commenters and comments, but this may only be a consequence of the selected channels and not a result of the behavior of the audience from the two countries.

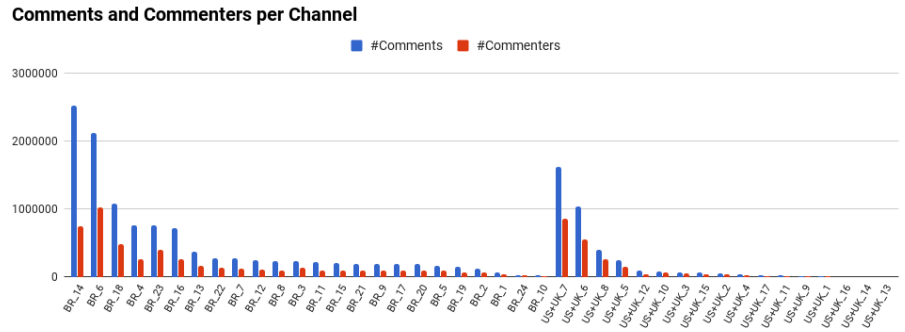

Fig. 3. Number of commenters and comments per channel.

Interestingly, the ranking of channels regarding the number of comments and the number of commenters is different. For instance, the channel with more comments in Brazil is BR_14, while the one with more commenters is BR_6. This result comes from the fact that the consumers of some channels are more active than others with respect to interacting (i.e., commenting) with the video.

In Figure 9 (Appendix) we present the number of comments and likes by video duration in seconds. From the chart we observe that the number of comments and likes does not correlate to the duration of the video, the most popular videos (i.e., videos that receive more likes and comments) are the shortest videos.

Figure 4 presents the cumulative distributions of four video metrics (number of visualizations, number of comments, duration in seconds and proportion of likes), comparing between the countries. We fitted the metrics to a Lognormal distribution and we estimate the parameters μ (mean) and σ (sd) using the maximum likelihood estimation technique. We present the corresponding function of distribution and the estimated parameters in the plots.

As we observe, the shape of the curves are similar between both datasets, although presenting different values. For instance, US and UK videos have more visualizations, while Brazil videos have more comments and a higher proportion of likes.

5.3 Analysis II: Advertising

According to Westenberg [19], YouTubers are viewed as authentic by their audience, when reviewing a product or brand. Followers believe that Youtubers' recommendations are honest. In order to look more honest and transparent to

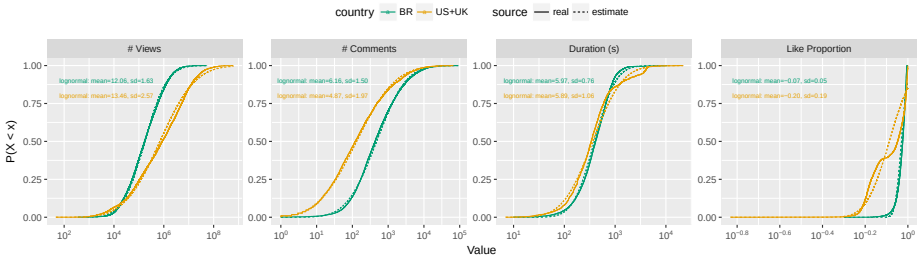

Fig. 4. Distribution of Video Statistics

their followers, Youtubers label their promoted videos with special hashtags, meaning the content, product or brand is sponsored. Thus, we take advantaged of the presence of hastags to identify the commercial nature of a video. Videos may contain explicit or implicit advertising. The former involves direct sales messages to a target audience. Implicit advertising, on the other hand, works best when businesses want to associate their brand or products with a psychological or symbolic element. We argue that if a video mentions products or brands, it potentially has advertising messages. To verify whether a video has advertisement, we employ the methodology of video classification explained in Section 4.4. We were able to classify 219 out of 1,055 tags for Brazil, and 249 out of 1,010 tags for channels in English. In total, 6,017 videos in Brazil were classified, and 4,109 videos in English.

Figure 5 shows the distribution of the video categories for both Brazil and English channels. The categories "Toys", "Movies and Shows" and "Games" are very popular in both countries, while "Services" is popular only in Brazil.

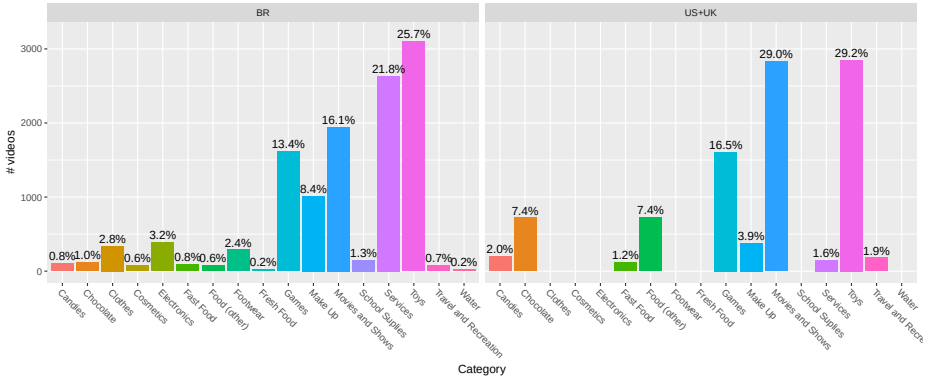

Fig. 5. Frequency of Video Categories by Country

Figure 6 presents the distribution of categories among all the 41 channels. We observe that there are some channels specialized in a certain kind of content. For instance, channels BR_15 and US+UK_11 have a high proportion of videos about Toys. Channels BR_6 and US+UK_7 are mostly about Games.

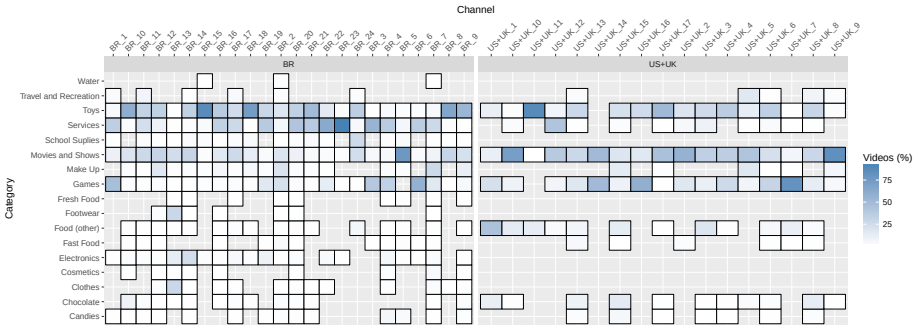

Fig. 6. Frequency of Video Categories by Channel

5.4 Analysis III: Audience

In this section our analysis focuses on YouTube audience. Considering only videos that have some kind of advertisement (i.e. those we were able to classify into one of the categories). The average number of comments per user is 3.19, for Brazilian videos, and 1.74, for English channels.

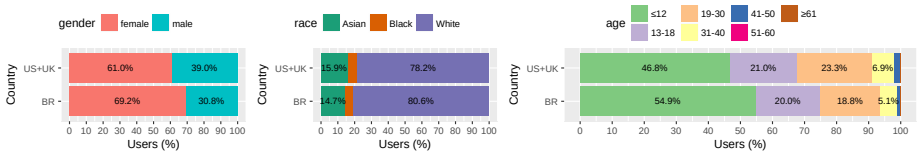

Fig. 7. Demographics of Audience by Country

First, we show gender, race and age distributions of commenters in Figure 7 - considering only videos with advertisement. The difference between men and women is smaller for users in English channels, but the proportion of women is higher in both countries. In the race distribution (middle) we observe a similar distribution for both countries showing that the proportion of white people commenting on videos is higher than the proportion of Asian and black people. On the bottom chart we observe that the user age distribution for both countries present similar behavior, the only difference is that in Brazilian videos the second age group who most commented on youtuber are children and teenagers between 13 and 18 years and, for the English channels, they are young people from 19 to 30 years. We choose these age intervals to distinguish from children under 12 years old, who should not have a YouTube account, and children and teenager audience between the ages of 13 and 18 who may have an account under parental supervision.

Figure 10 (Appendix) shows the demographics with respect to gender, race and age groups for each channel. We observe that the profile of the audience might be very different. For instance, 90% of channel US+UK_1's audience is female, while this is only 52% for channel BR_3. Regarding age, some channels have a child audience of nearly 70%, such as channels US+UK_15 and BR_10.

Figure 8 shows the demographics for each category, for both countries. We observe that the Games category is the one with higher proportion of male audience, although still having more women comparatively. Also, the audience for Games is older, presenting lower frequencies of children.

Brazilian legal framework considers publicity aimed at children as abusive and, therefore, illegal. The legal definition of children includes any person under 12 years, and the Consumer Code, which specifies a set of abusive conducts, lists as one of them the act of directly approach children with publicity of products or services, considering they haven't reached a certain degree of bio-physical development which is necessary to identify and understand the marketing discourse and, therefore, is legally regarded as vulnerable. Brazilian law regarding Internet - basically the Internet Civil Rights Framework, or Marco Civil da Internet - basically follows this same rationale and recognizes the need for special information and education about children's access to Internet. There is no general data protection framework enacted in Brazil which could impact children's privacy.

The collection and use of Children's personal data is subject to the standards of COPPA (Children's Online Privacy Protection Act of 1998), which dictates that no data regarding persons under 13 years can be collected without their parents or caregivers explicit consent. COPPA also includes a series of obligations for site owners, makes it mandatory for a website to include in its privacy policy a set of rules and warranties for the its usage by children, and also clarifies how the consent from the parent or responsible has to be collected.

6 Conclusions

Google has some clear age policies for its products[13]. The minimum age requirements to own a Google account in the United States is 13 or older (i.e., except for Google Accounts created in Family Link for kids under 13), 14 or older in Spain and South Korea, 16 or older in Netherlands and 13 or older in all other countries. Some services have specific rules, such as YouTube that specifies that age-restricted video should be watched only by users who are 18 or older. In Brazil, however, the use of YouTube itself is restricted to those over 18, according to the terms of service of the platform that is clear in stating that the YouTube website is not designed for young people under 18 years[14].

Among outcomes of this paper we could also mention that it can lead to a discussion about Google's politics on the age limit if it is confirmed the active presence of under 13 on YouTube. Also, data gathered and analyzed about the profile of users under 13 may be used in future research about (i) the presence of racial and gender bias, (ii) the means publicity approaches children on YouTube and (iii) the way private data from children is collected and commercialized in digital media.

Out of the data analyzed, we believe the major impact may result from the identification and characterization of children actively using YouTube. Even if

[13] support.google.com/accounts/answer/1350409?hl=en

[14] www.youtube.com/static?gl=BR&template=terms&hl=pt

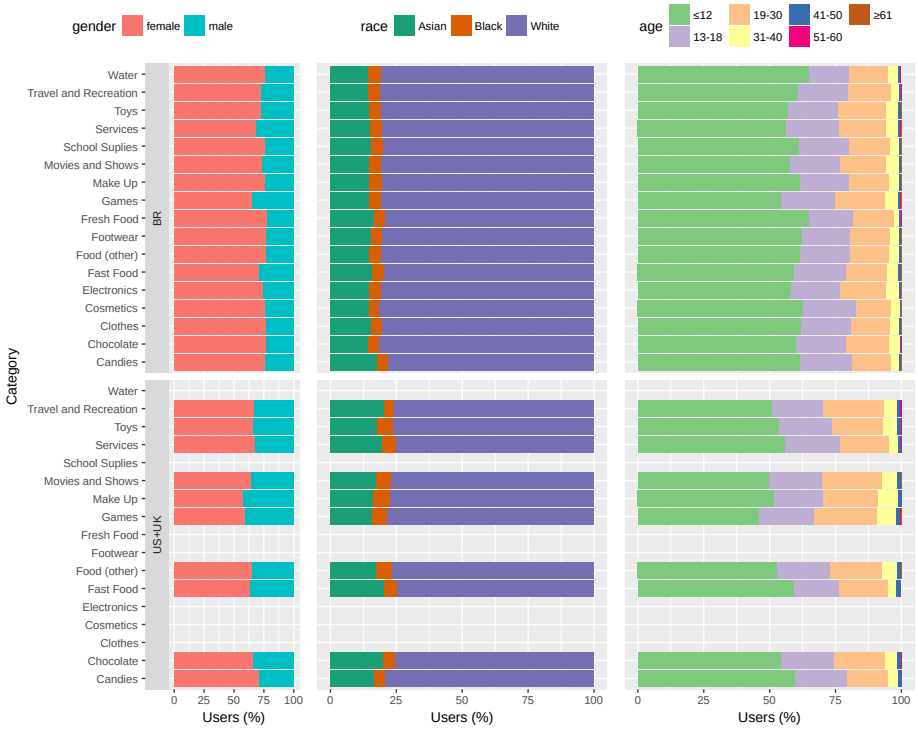

Fig. 8. Demographics of Audience by Category

some usage of under-18 is generally considered as fair due to parents' or legal responsible consent and supervision, the fact is that if children are actually using the platform they can be exposed to different challenges, as advertising, inappropriate content, privacy issues and crimes in the digital world, which raise concerns about compliance with regulations in several countries.

Other questions addressed by this work could be investigated in greater detail to highlight possible nuances not captured by the experiments done. For example, an evaluation of how exactly Face++ accuracy is impacted by the particularities of the pictures in the user profile could help to know whether there are adjustments to be made in this respect. A detailed analysis of usage patterns and spread of YouTube channels across countries may reveal how local differences affect the overall temporal dynamics found. Analysis of the influence of geographic and cultural location on the user behavior would be interesting for promoting educational and healthy food videos among children. Considering that many videos blur the boundaries between entertainment and advertising [6], another possible direction would be to dig further into the transcripts of the videos to analyze the texts and characterize the different types of advertising that are exhibited to children.

Acknowledgements

This work was partially supported by CNPq, CAPES, FAPEMIG, and the projects InWeb, MASWEB, and INCT-Cyber.

References

1. S. Bakhshi, D. A. Shamma, and E. Gilbert. Faces engage us: Photos with faces attract more likes and comments on instagram. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 965–974. ACM, 2014.
2. F. Benevenuto, A. C. M. Pereira, T. Rodrigues, V. A. F. Almeida, J. M. Almeida, and M. A. Gonçalves. Characterization and analysis of user profiles in online video sharing systems. *JIDM*, 1(2):261–276, 2010.
3. F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida. Characterizing user behavior in online social networks. In *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement*, IMC '09, pages 49–62, New York, NY, USA, 2009. ACM.
4. J. Chester. How youtube, big data and big brands mean trouble for kids and parents. `http://www.alternet.org/media/how-youtube-big-data-and-big-brands-mean-trouble-kids-and-parents`, 2015. Accessed: 2017-06-13.
5. M. Dehghani, M. K. Niaki, I. Ramezani, and R. Sali. Evaluating the influence of youtube advertising for attraction of young customers. *Computers in Human Behavior*, 59:165 – 172, 2016.
6. S. Dredge. Why youtube is the new children's tv... and why it matters. `https://www.theguardian.com/technology/2015/nov/19/youtube-is-the-new-childrens-tv-heres-why-that-matters`, 2015. Accessed: 2017-03-15.
7. S. Dredge. Joseph garrett, the children's presenter with 7.8 million subscribers. `https://www.theguardian.com/technology/2016/aug/28/stampy-joseph-garrett-youtube-childrens-presenter-millions-of-viewers`, 2016. Accessed: 2017-06-13.
8. S. Livingstone. Children's digital rights: a priority. *Intermedia*, 42(4/5):20–24, 2014.
9. S. Livingstone. Digital media and children's rights. `http://blogs.lse.ac.uk/mediapolicyproject/2014/09/12/sonia-livingstone-digital-media-and-childrens-rights/`, 2014. Accessed: 2017-03-15.
10. S. Livingstone and C. Local. Measurement matters: difficulties in defining and measuring children's television viewing in a changing media landscape. *Media International Australia*, 163(1):67–76, 2017.
11. F. Lupianez-Villanueva, G. Gaskell, G. Veltri, A. Theben, F. Folkford, L. Bonatti, F. Bogliacino, L. Fernandez, E. Marek, and C. Codagnone. Study on the impact of marketing through social media, online games and mobile applications on children's behaviour. `http://ec.europa.eu/consumers/consumer_evidence/behavioural_research/impact_media_marketing_study/index_en.htm`, 2016. Accessed: 2017-03-15.
12. G. Magno, G. Comarela, D. Saez-Trumper, M. Cha, and V. Almeida. New kid on the block: Exploring the google+ social graph. In *Proceedings of the 2012 Internet Measurement Conference*, IMC '12, pages 159–170, New York, NY, USA, 2012. ACM.

13. G. S. O'Keeffe and K. Clarke-Pearson. The impact of social media on children, adolescents, and families. *Pediatrics*, 127(4):800–804, 2011.
14. G. C. on Internet Governance. One internet. `https://www.ourinternet.org/sites/default/files/inline-files/GCIG\_Final%20Report%20-%20USB.pdf`, 2016. Accessed: 2017-03-15.
15. R. Ottoni, J. P. Pesce, D. B. L. Casas, G. F. Jr., W. M. Jr., P. Kumaraguru, and V. A. F. Almeida. Ladies first: Analyzing gender roles and behaviors in pinterest. In *ICWSM*. The AAAI Press, 2013.
16. J. W. Pennebaker, C. K. Chung, M. Ireland, A. Gonzales, and R. J. Booth. The Development and Psychometric Properties of LIWC2007. This article is published by LIWC Inc, Austin, Texas 78703 USA in conjunction with the LIWC2007 software program.
17. A. K. Przybylski and V. Nash. Internet Filtering Technology and Aversive Online Experiences in Adolescents. *The Journal of Pediatrics*, Mar. 2017.
18. M. Stoilova, S. Livingstone, and D. Kardefelt-Winther. Global kids online: Researching children's rights globally in the digital age. *Global Studies of Childhood*, 6(4):455–466, 2016.
19. W. Westenberg. The influence of youtubers on teenagers: a descriptive research about the role youtubers play in the life of their teenage viewers. MSc dissertation, University of Twente, September 2016.

A Appendix 1

Dataset Fields		
Video	channel id	Video id
	channel name	Id of channel where video was posted
	video id	Video channel name
	video title	Title of the video
	video description	Video description (made manually by youtuber)
	transcript	Automatic textual transcription from the audio
	subtitle	Manual subtitle (made by youtuber or by third parties)
	video tags	List of tags (made manually by youtuber)
	video date	Video posting date and time
	video duration	Video duration in seconds
	view count	Number of views
	comment count	Number of comments
	like count	Number of likes
Comment	comment id	Comment id
	author name	Name of the commenter
	author id	Id of the commenter
	author image	YouTube profile picture of the commenter
	comment date	Date and time the comment was posted
	comment text	Content of the comment
	video id	Id of the video (in which the comment was posted)
	parent id	Id of the original comment (if it is a comment reply)
User	gender, age and race	features extracted by Face++

Table 3. Dataset Fields

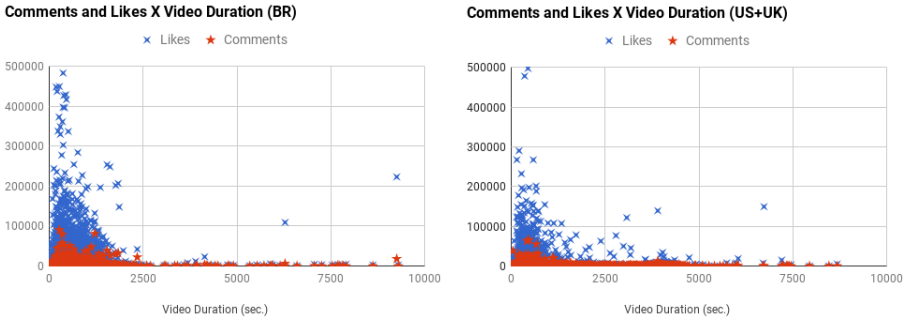

Fig. 9. Comments and Likes X Video Duration.

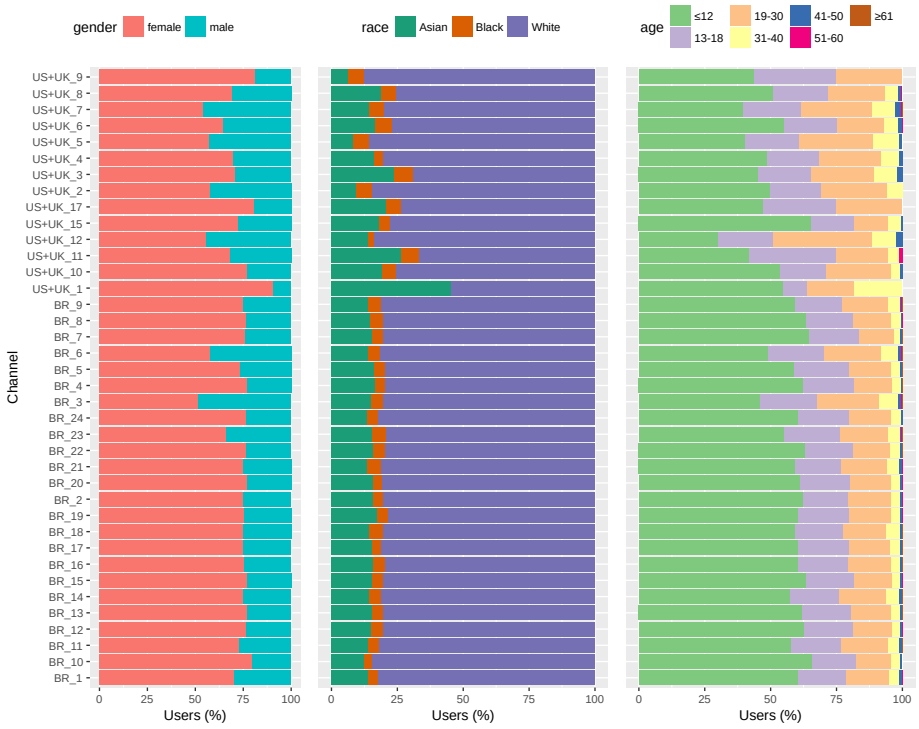

Fig. 10. Demographics of Audience by Channel

B Appendix 2 - Semantics

In this analysis we want to measure how the video was evaluated by the viewers. We use the text of the comments as a proxy for the perception of the audience, looking into the semantics it contains. We focus on only two categories of LIWC:

Positive Emotions and Negative Emotions. Since the LIWC is available only for the English language, we inspect only comments from the U.S. channels.

Figure 11 (Appendix) presents the percentage of the comments that contain words related to positive emotions or negative emotions, according to LIWC. We aggregate the comments by channel, video category, gender and age group. The predominance of positive emotions is notorious, indicating that the videos are, in general, well evaluated by the public. Interestingly, some channels have a higher proportion of positive words than others, such as US+UK_11 and US+UK_9. Regarding the video categories, we observe that videos with make up are more positive than the others. Looking into the social groups, there are no huge differences, although we observe an indication that the use of positive words seems to decrease as the audience get older.

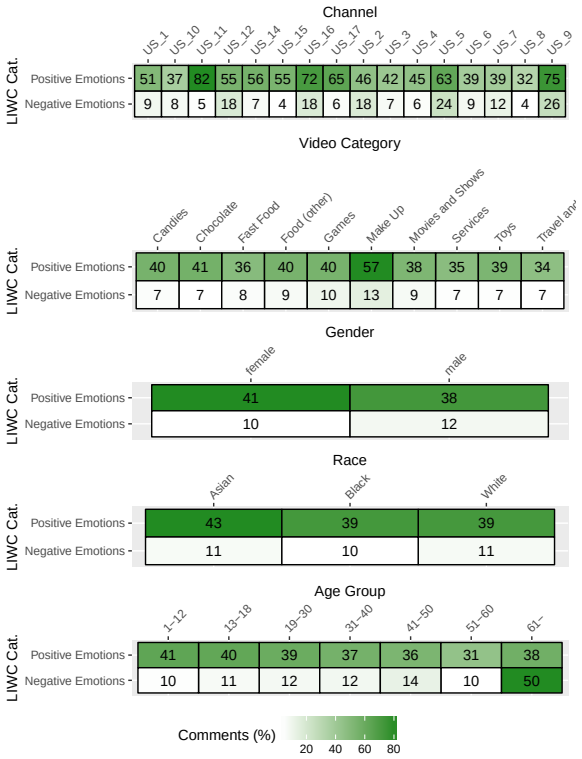

Fig. 11. Percentage of Comments Containing words of Positive and Negative Emotion