Pornhub Tracking Exposed

Researching a method for the critical and independent comprehension of the algorithmic personalization in pornography –the Pornhub case.

1. Introduction

Advancing the critical analysis of the online pornographic phenomenon can bypass an approach content-based only, to focus instead, and through a transdisciplinary approach, on the effects that the data economy has on the porn industry. This research includes a historical perspective that aims at understanding the role of cyberporn in the development of the Internet while taking into account critical studies on mainstream technologies. It has been noted that a tool is required to further the analysis and increase public awareness regarding digital platforms' concerning aspects. In the domain of porn sites, Pornhub is the biggest and the most important actor for content personalization. Furthermore, it utilizes data analysis as a plus value and as an important communication and marketing tool[1], in a manner of speaking it does also a data pornography. The research methodology simulates different online identities, this

---

[1] https://www.pornhub.com/insights

is functional to verify the research hypothesis and increases the possible scenarios through a design that keeps in mind the alterity. Pornhub, like its competitors and other popular social media, collects a certain amount of data on how users interact with the platform. These data, besides being useful for market analysis, are also used to customize the user experience, even if sometimes the users are unaware of it. On the homepage there is a section that proposes to the users videos that should be satisfying for them, according to the company and thanks to the personalization algorithm, based on what they saw before and how they saw it. Pornhub processes the data in a way that is unknown to the users and then somehow takes for granted or knows how to predict their tastes and then make a profit from it. pornhub.tracking.exposed is a browser extension that allows to study ed analyze the implications of this personalization algorithm through the collection and comparison of individualized experiences that Pornhub users receive.

2. The role of cyberporn in the development of the Internet

It is difficult to gather complete and verified historical information about the cyberporn phenomenon; surprisingly no one has ever written a complete historiography of the pornographic industry on the internet. Data and all other materials are difficult to combine together and it seems that the topic is not considered legitimate for the history of business or communication. Even the porn industry itself is often a little shy when it comes to shedding light on its marketing practices and strategies.

It is commonly recognized that for all new technologies and media porn is moving towards a subsequent adoption by the mainstream public. One of the first artifacts that were produced after the invention of printing was certainly the Bible, but along with it there were licentious and obscene poems and stories. Concerning the photographic trade at the beginning it was mainly about selling portraits to people, of themselves or their loved ones, but the real turning point for the photographic market was the trade in photos of other naked people. These photos were labeled as study models for young artists. The soldiers in Crimea in 1850 usually carried pictures of their loved ones, but

also these so-called "study models". The same process happened for radio, cinema, television, everything has been pornized.[2]

Regarding the history of computers, for a long time porn had no role in its development, in fact at the beginning it was simply because PCs couldn't yet manage images. Although there was still a private exchange of risqué material among ARPANET researchers and scientists, there wasn't still a way to distribute and consume pornography on a large scale. Since the late '80s the Internet has become almost synonymous with pornography in the public imagination and it can't be denied that porn has played an important role in popularizing the web. Some technological developments of the '80s have made it possible: the VGA graphics have finally allowed the display of images, hard drives have allowed storage in personal computers, networks then allowed sharing, trading and peer to peer.

The pornographic images became an easily digitizable content in which people were very interested, willing to wait patiently, as well as willing to pay and use their credit cards to do so. The Usenet protocol, after the introduction of the ASCII character encoding, made it possible for the first pornographic images to be created with text.

In the early '90s, Bulletin Board Systems or BBS began to expand. The vast majority of these BBSs were small and private. But some system operators, called Sysops, had understood that porn images had popped up among the most popular requests and, coinciding with the appearance of the first digital scanners, they realized that by scanning Playboy or Penthouse they could quickly create a library of images for which people would have paid a lot. Some of them began to charge access to their systems. Event Horizons BBS grossed over $3.2 million in the year 1993[3], employing ten people to scan photographs and put them online for download. In order not to do all the work on their own, the BBS began to talk about compensation systems and to grant permissions to download 4 bytes for each byte uploaded, so that users were led to upload and share their personal collections. From that moment every user could connect to a BBS and have access to more porn content than had ever been available anywhere ever before.

---

[2] Soldiers enjoyed unprecedented access to obscene materials of all sorts. See also:Giesberg, Judith Sex and the Civil War: Soldiers, Pornography, and the Making of American Morality. 2017, University of North Carolina Press.

[3] https://www.wired.com/1993/04/bbs/

When producers of the likes of Playboy noticed it, the first complaints of copyright infringement appeared in the digital environment, when images could yet not be seen directly on the web.

At the climax of the BBS industry, there were more than 45,000 in the United States, and although most of them were free, those who had paid subscriptions had a turnover of 100 million dollars a year, in an era in which almost no one was online: all thanks to porn. In addition to the exchange of images there were obviously also exchanges of words and conversations between users. The message boards of the BBS systems were full of adult forums where users could have telematic sex. Internet Relay Chat (IRC) was developed on a Finnish BBS in 1988[4] to replace the previous BBS chat clients in real time and the same happened for AOL, where users paid to enter private chat rooms to be able start sexting and exchange anonymous files. These were the premises for the birth of the World Wide Web, the first computer technology that as a main component of its value proposition had the mainstream adoption of pornography.

The first porn sites were transfers of collections of BBSs on the web, but it was immediately clear that scanning images would not be the future, thus some began to produce their own content, encouraged by the low costs of production and the great possibility of earnings. Furthermore many contents from Eastern Europe, with the fall of the Berlin Wall, could be bought at low prices and sold to Americans.

Already from the second half of the '90s, cyberporn began to be studied as a cultural phenomenon in the academic sphere, not without controversy. One of the most striking examples, with a particular media implication, was Martin Rimm's study at Carnegie Mellon University, conducted mainly on the first computer porn appeared in Bulletin Board Systems, starting with a selection of 917,410 sexually explicit files from BBSs for adults[5]. The TIME, in June 1995, completely misrepresented the study, contributing to the demonization of the internet as a virtual place for meetings between disreputable people; this stating that 83.5% of the contents present on the Internet were pornographic, when the research was conducted on BBS for adults, was wrong and, therefore, the presence of pornography was rather obvious.

---

[4] http://www.irc.org/history.html

[5] https://jmetz.com/2015/03/misunderstanding-cyberculture-martin-rimm-and-the-cyberporn-study/

"CYBERPORN EXCLUSIVE: A new study shows how pervasive and wild it really is. Can we protect our kids - and free speech?" This is the title of the article which fortunately was followed by much criticism both by academics, for the unclear research methodology and the anti-scientific approach of the article; from sociologists for the provocative thesis that the cyberporn was guided by a request for images that could not be found in newspapers (pedophilia, sadomasochism, pissing, scat and bestiality); from those who saw it as an attack on freedom of speech; from the virtual community itself, which has had to contend for years with this prejudice that sees the Internet as a meeting place for perverts. The first to react to these attacks were two professors from Vanderbilt University, Donna Hoffman and Tom Novak, who directly accused from their site Rimm's study and the article in the Time of containing serious conceptual, logical and methodological errors. Martin Rimm disappeared completely, but the article was published at a very critical moment in the evolution of the Internet, when the US Court was about to discuss the Communication Decency Amendment, a decree law of censorship against material considered obscene on the net, which aimed above all to protect minors and which passed with 84 votes against 16, but was later declared unconstitutional, as in contrast with the First Amendment[6].

Porn was also fundamental for the creation of technological systems such as the creation of an account, the verification of it and security in online payments and therefore in the development of digital marketing. It is now common practice for any website or app to register a new user with a double activation process. You sign up with an email address and password, but then the site or app sends you an email with a link to click on before having the full privileges on the site. This practice of double-opt-in signup ensures that a person is real and reduces the creation of spam accounts; it was conceived and tested by Cybererotica[7], also to avoid that a person, victim of a prank, received unwanted emails with pornographic content. Early pornographers understood this could be an excellent marketing practice, but above all a way to avoid being included in the black list of ISP spam lists like AOL.
Cybererotica was even more innovative. Ron Levi, its owner, invented a model of affiliate marketing, which Amazon has adapted to its business and now holds the patent.

---

[6] https://internetlaw.uslegal.com/pornography/

[7] Lane, Frederick S. Obscene Profits: Entrepreneurs of Pornography in the Cyber Age, 2001, Routledge.

FastCash, this was the name of Levi's PayPerClick affiliate program[8], allowed affiliated webmasters to receive small amounts of money for each user click on mini porn sites that opened along with their web pages, which showed free examples of pornography, referring to the largest paid site. Web advertising did not start until the end of 1994, and even when banner advertisement was particularly popular, mainstream sites did not want to put any pornographic ads on their pages, so Levi decided that the best way to advertise was to make many other pages appear and hope that someone was attracted to them.

Levi also developed, in 1996, one of the first web data analysis tools: XXX Counter, to help webmasters better count unique visitors compared to simple clicks; he sensed that it wouldn't have been better to pay for each click, but rather based on the results. So from 1997 Levi first moved to Pay Per Click and finally to Pay per Action, that is he paid the affiliates only if a user had actually signed up. Pay Per Click and Pay Per Acquisition were not adopted by the mainstream advertising industry until the early 2000s. Because of the puritanical face of the mainstream internet, porn had been largely excluded from traditional marketing, often developing in a separate universe, with business models that changed and advanced at exponential speed compared to the traditional market, on an anarcho-liberal model.

Another important innovation attributed to Levi is the development of modern live video feeds in May 1997. Clearly, the video was very important for internet pornographers, but providing minimal quality streaming video had been a difficult technical proposal in the time of dial-up. Users were enthusiastic about video codecs and plug-ins such as Microsoft's Real Networks and Windows Media Player, but digital pornographers soon found out that downloading and plug-ins made the porn experience more difficult and puzzling. The pioneer of video streaming was the Dutch company Red Light District, fun.nl, which developed a compressed video streaming system as early as 1994[9]. Other solutions developed on an industrial level then allowed to reproduce heavily compressed porn videos, easily downloadable directly from the web pages, without the need for plug-ins. Even with 28K modems, these video streams allowed users to watch

---

[8] Waskul D. Net.SeXXX: Readings on Sex, Pornography, and the Internet, 2004, Peter Lang

[9] Barss, P. The Erotic Engine: How Pornography Has Powered Mass Communication, from Gutenberg to Google, 2010, Doubleday Canada

videos by starting or stopping at their leisure, with the ability to skip back and forth in the timeline, a feature that then YouTube really liked.

Already tens of thousands of full-length adult film titles were available on the web when, around the 2000s, Hollywood began its digital attempts to commercialize their videos and films. Another sector in which porn excelled globally was the one of videoconferencing, based on the business model of the paid erotic phone call. In 1997 the first live cams were born, where thousands of cam girls perform: they are live performers in front of webcams, hosted by sites paid by viewers.[10] In a moment in which hardware companies were struggling to convince large companies to adopt video conferencing, the adult industry was enjoying it on a large global scale. Just as blogs and social media allowed anyone to have a voice, the web had already allowed anyone to be a porn star.

Before Web 2.0 and social media, before the YouTube stars and Facebook accounts, it was porn that led net surfers to the concept of user-created content and made the maximum "Content is King" internalized in users' minds, even before 1996, when Bill Gates said it[11]. Homemade pornography was made possible by the technological advances of cameras, video cameras and mobile phones, but it's the web that has allowed its distribution in an endemic way.

3. Cyberporn and data economy in web 2.0

YouTube in 2005 introduced into the mainstream the experience of freely uploading content, watching and sharing it directly on the browser. It was meant for User Generated Content, but rapidly was populated of multiple uploads of professional videos, TV show ripped, copyrighted titles, and music videos[12]. Similarly, the porn version of the technology, followed the same patch. Now known as Tube Sites, (Pornhub, YouPorn, RedTube, and TubeGalore) these were first advertised as a place where people could share their personal videos, like homemade porn. As you can expect, the ripping phenomenon replicated immediately and not just from existing porn DVDs, but also of

---

[10] https://en.wikipedia.org/wiki/Webcam_model

[11] https://medium.com/@HeathEvans/content-is-king-essay-by-bill-gates-1996-df74552f80d9

[12] https://en.wikipedia.org/wiki/History_of_YouTube

material freed from the password-protected firewalls of the paid porn sites. Compared to YouTube dynamics, the producers, found initially tricky to react legally.

According to the Wall Street Journal[13], between 2008 and 2010, adult TV pay-per-view revenue fell from $1 billion to $899 million. DVD sales and rentals fell by 22% in 2008 alone, twice as much as in Hollywood. Steven Hirsch, head of the Vivid Video studio, estimated that overall DVD sales had dropped by 80% in the five years prior to 2014.

The Tube Sites that have hindered the porn industry so much can hide behind a clause in the Digital Millennium Copyright Act. This clause states that site owners are not responsible for anything their users do while enjoying their services: thus, declaring that the contents protected by copyright are loaded by the users, they are protected from any legal implication. According to an article[14] in New York Magazine in 2011, a manufacturer of pornographic material called Brazzers has actually started some of the biggest Tube Sites: they have used these sites to promote their content, in coexistence with other producers' content, uploaded by users or not. Brazzers then sold it to a German programmer named Fabian Thylmann, who changed the name of the general holding company to Manwin, based in Canada and Luxembourg.

Currently, the porn industry seems to have found a way to coexist with Tube Sites, which have actually taken control of it and which economically now have less to do with video and more with user data. Porn has gone from being an unmissable, transgressive, rare proposal, with a quality standard that people would pay for, to a kind of trivial content, at no cost, as well as almost any other digital asset.

The frustrations of the pornographers were exactly the same as their big Hollywood brethren. The porn producer Scott Coffman said: "I can contact a tube site and tell them to take my contact down, and it's their responsibility to do so. But, if someone in the next hour uploads [the same material] to the site, then I have to call again and have the same conversation. [The pirates] can have one person sitting there all day continually uploading content, and [the studios] are expected to police this activity. **It would be more effective if the owner of the tube site was responsible for policing it**. In pawn

---

[13] https://www.wsj.com/articles/SB10001424053111903885604576488540447354036

[14] http://nymag.com/news/features/70985/

stores, for example, there is a responsibility not to resell stolen products. If the DMCA worked the same way, it would be a different story for the industry."[15] Now, as stated by Pornhub, it seems that thanks to artificial intelligence, machine learning and face recognition this type of problem can be solved and this will also allow to automatically catalog pornstars and videos through automatic tags (even if to process all the videos may take a few years). Pornhub vice president Corey Price said in 2017: "Ultimately, we want to provide our fans with whatever it is they have a penchant for, and our new model will be able to surface more accurate results for them, which will hopefully keep them coming back for more, the easier it is to navigate and access what they want, the happier they will be."[16]

We have to take a step back in time, in the last decade. Internet was even less monitored than now, and this made the sharing of copyright content very common t. These companies were producing revenue with profiling, advertising (not uncommon, even spreading malware or gambling ads). It was a different distribution model compared to the previous one, based on VHS and DVD, and this is the key to interpret the events we are observing now. The peer to peer is a distribution network invulnerable to these risks, but somehow, downloading an application and becoming an expert in finding trustworthy sources was harder than googling for a porn website. Web browser establish themselves as the tool to consume online porn when flash and java applets permit the immediate rendering of the video in web pages.

The "official" protection framework is the DMCA (Digital Millennium Copyright Act) which shifts the effort on the side of the copyright owner. The complainer sends a takedown request, and the platform begins to delete content. This approach failed in porn (and in non-porn) platforms for different background reasons.

In porn, companies have video material, and some of them were copyright-owned. If the titles, actor names, content length might differ to make the analysis harder for the copyright owner. For example: if I upload an X-Men movie on youtube, Marvel copyright owner will keep searching for x-men and other keywords, such as super-hero names. In

---

[15] http://www.internethistorypodcast.com/2015/01/history-of-internet-porn/

[16] https://venturebeat.com/2017/10/11/pornhub-taps-computer-vision-to-identify-porn-stars-and-automate-content-tagging/

a porn website this might be harder: the title may or may not have the name of the actress involved. It might just describe body features. Only a video analysis might reveal the violation, but this can be hardly automated now, and it was impossible in the past. Also, the Internet of the previous decade was much less regulated, surveilled, and was used by entrepreneurs with dodgy methods. You should not think of an established company such as Google/Youtube (which answer at their best to DMCA and regional copyright infringement). A good chunk of these porn websites were fringe services which might try to escape their responsibility. The use of offshore services, anonymous registration, and spread malware when client-side exploitation was not uncommon. They were benefitting of the stigma on porn secrecy, to make it less likely to be reported by customers for unfair behavior. Distributing copyright content under premium access was likely the least sketchy business for such kind of company.

In the non-porn industry, the reader might start by reading the debate of 2018-2019 on the European Copyright Directive. The so-called *upload filter*[17], represents an essential shift of responsibility (this is not yet implemented law. It is a "directive", and the European member State would create national laws with individual differences).
The previous law you think the leading law before was making the ISP not responsible for the user-generated content[18]. This, of course, assumes good-faith service providers.[19]
In an April interview[20], Speaker Nancy Pelosi of California called Section 230 a "gift" to tech companies "that could be removed."

One of the golden rules of the Internet is to try to grow as much as possible at the beginning, "move fast and break things" kind of approach. When the company becomes the biggest actor on the field it will start to respect regulation, ask for self-regulation, kill competition, and change attitude. This current situation can be an explanation to PornHub's announcement of using Artificial Intelligence to recognize actress faces, copyright scene, and implement in this way their upload filter.

---

[17] https://www.theverge.com/2019/3/26/18280726/europe-copyright-directive

[18] https://www.eff.org/issues/cda230

[19] https://www.nytimes.com/2019/08/06/technology/section-230-hate-speech.html

[20] https://www.vox.com/2019/4/12/18307957/nancy-pelosi-donald-trump-twitter-tweet-cheap-freak-presidency-kara-swisher-decode-podcast-interview

But even using artificial intelligence for this type of purpose cannot be immune to criticism. A 2016 study[21] by Robert W. Gehl, Lucas Moyer-Horner and Sara K. Yeo of the University of Utah critically examines filtering of artificial vision-based pornography (CVPF), a subfield of computing that seeks to train computers on how to recognize the difference between digital porn graphic images and non-pornographic images, which doesn't mean recognizing the face of actresses but distinguishing between different types of content, for example not relevant to the site.

Based on 102 peer-reviewed articles, the study claims that computers are trained to "see" a very specific and idealized form of pornography: mostly images of solo, thin and naked women. The breast exactly twice the size of the head, the hips one and half that size . Here is the ideal form of pornography. It may not be the ideal form for most people interested in online porn, but it's ideal for a specific type of consumer: the computer trained to recognize pornographic images.

The idealized pornography for artificial vision is symptomatic of how computer scientists are transmitting their ideas into their technological artifacts. They make, transmit and contribute to the dissemination of arbitrary assumptions about sexuality, pornography, gender and bodies. Sexuality is not a man who looks at a naked woman and pornographic bodies are not specific or predictable sizes, shapes and proportions. The study carefully considers the algorithms proposed in the CVPF literature and quantitatively analyzes the images considered with these algorithms, demonstrating that the algorithms are trained with a training test that contains more images of women than men. Based on feminist and porn studies, the paper also compares the CVPF pornographic imagination with noise-pornography that goes beyond artificial vision, but which would instead be representative of the heterogeneity of networked pornography.

Now that the first generation of kids who have never known a world without cyberporn is becoming of age, many have wondered how big the pornographic industry has actually become. The web is full of data, even conflicting, but the question is: who should we believe? Certainly it's extremely easy to access cyberpornography and video sharing services, such as Pornhub, which are among the 15 most visited sites in European countries (according to alexa.com), and among the top bandwidth consumers in the

---

[21] https://journals.sagepub.com/doi/full/10.1177/1527476416680453

world[22], while not directly producing any kind of pornographic material. This underlines its importance as a gatekeeper, as an intermediary, therefore important to study since we should never assume that a technology is politically neutral. The problem is not that there is a gatekeeper, or even that this is not human, the problem is that gatekeeping takes place behind closed doors. There is no supervision, no instance of control, considering that there is no objective machine, as there were no objective newspaper publishers. The statistics released annually by the site, the Pornhub Insights, can help to understand the magnitude of the phenomenon. It's clear that it is no longer appropriate to think about porn as a niche genre and not even as a specific narrative genre, as it's "overflowing from the subcultural margins in which it has crouched until the explosion of the participatory version of the Internet" (Attimonelli and Susca 2016) and allowed the advent of widespread pornculture, which went very well with pop culture. Between porn and the rest of the communication system there are no longer barriers as before, when to get sexually explicit material a certain embarrassment had to be overcome; now porn permeates the digital everyday life and beyond. The same increasingly private and individualistic nature of the devices could also derive from the need for private consumption of pornography: no more discomfort in buying a magazine, in renting a DVD, in sitting in the audience in a porn cinema, in using the TV or family computer. The Internet and its devices could be understood as the climax of pornographic technology and not the opposite. In order to prevent children from accessing it, it's become common habit to use Parental control on devices owned by kids, as well as network traffic classification and website blacklisting in corporate networks (which is not always effective[23]).

Pornhub's business model was also initially able to take advantage and promote UGCs, user generated contents. This has completely changed the offer, contributing to the rise of the amateur category. The gratuitousness and the amateur production are constant elements in the historical experience of Pornhub, while up to that moment the pornographic industry was closely connected to publishers, authors and actors. Along the lines of social networks, which introduced us to concepts such as *personal branding*,

---

22

https://slate.com/technology/2014/10/mindgeek-porn-monopoly-its-dominance-is-a-cautionary-tale-for-other-industries.html

23

https://www.cnbc.com/2019/07/23/facebooks-messenger-kids-app-flaw-let-users-chat-with-strangers.html

potential producers have increased exponentially. The Verified Amateurs category strengthens the *coherence* of the brand, which in this way declares that knows how to distinguish a film of amateur-style professionals, from videos that are truly amateur. The UGC is therefore an original *value* but also the way in which users become loyal, so the user "becomes both a producer and a consumer at the same time as the platform he helps to develop" (Ciuffoli, D'Amico and Pulzelli 2018). The desire to build a porn community is what most distinguished Pornhub compared to its competitors. Proof of this is the inclusive business model through the Model Payment Program and a social network dedicated to members of the Pornhub Network.

From 2013 the brand's efforts begin to make the users even more protagonists, they become the reference point for the brand itself. Videos take on a more marginal role and become "the means to a greater goal that prepares the company for the development of very different strategic areas" (Ciuffoli, D'Amico and Pulzelli 2018). One of the user-centered design strategies implemented was the opening of Pornhub Insights, a blog that collects, analyzes and shows with engaging graphics the behavior of users who browse the site. Clear that the user is indispensable as it is the content itself, through his data, for which Pornhub is committed to understanding his needs. This blog is the manifestation of a brand-oriented, and consequently business-oriented, idea that bets on porn as a mainstream product, which wants to open up to the general public to *enjoy* all the associated market implications. Once the time of the numerous and fragmented porn communities is over, it will be time to move towards the universally recognized sex community.

In addition to having a sex community on the site, sensitive data on site visits and therefore onanistic acts have now become, for companies like google and facebook, one of the parameters that define the uniqueness of a user, and somehow also characterize an entire community.

The study "Tracking sex: The implications of widespread sexual data leakage and tracking on porn websites"[24] by Elena Maris, Timothy Libert, Jennifer Henrichsen found that Google (or one of its subsidiary companies like the advertising platform DoubleClick) had trackers on 74 percent of the pornography sites. Trackers from the

---

[24] https://arxiv.org/abs/1907.06520

software company Oracle showed up on 24 percent of sites, and Facebook, which does not permit[25] pornographic content or nudity on any of its platforms, had trackers on 10 percent of the sex websites scanned by the study. Where that data ultimately goes is not always clear. The New York Times highlights the ethically most crucial point of the issue:

> Dr. Maris argues that this lack of disclosure is similar to the issue of sexual consent. "As in any sexual interaction, silence must not be mistaken for consent," she said. "Individuals should have a clear understanding of the power dynamics of the sexual exchange they are entering when visiting porn sites." Those power dynamics, according to Dr. Maris, are deeply unbalanced. "You have some of the world's most powerful companies here," she said, noting that there's very little redress for the consumer should the data end up in the wrong hands.
>
> Affirmative consent is at the heart of digital privacy. Nearly all tracking is by default and governed by impossible-to-read privacy policies. And in an era that privileges and prioritizes mass collection of personal information, that means gathering information that is not only invasive but also superfluous. The leaky user data of pornographic websites is merely an extreme example of what has become standard practice online."[26]

The GDPR represents a first European position in the comparison of data processing and is therefore a response to unbridled tracking of the last fifteen years. In the embryonic stage, on the other hand, the study, monitoring and understanding of the implications of algorithmic personalization, aimed at its possible regulation to defend civil rights and user and society's well being. The control over what should be profitable and the limits to the exploitation of data have not been made public yet, but remain a privilege of those who collect, own and analyze the data.

To understand the importance of privacy in prostitution, whether filmed or not, it's possible to reflect on the fact that in October 2017 Gizmodo accused Facebook[27], because with its operations it would jeopardize the identity of sex workers, exposing their true identity to customers among the friendship suggestions. A California prostitute, who tried to keep her two identities separate, found some of her regulars clients among the "people you may know" of her personal account. The fact that she

---

[25] https://www.facebook.com/help/community/question/?id=107190583098138

[26] https://www.nytimes.com/2019/07/17/opinion/google-facebook-sex-websites.html

[27] https://gizmodo.com/how-facebook-outs-sex-workers-1818861596

used a different name, phone and email address was not enough. In this case, in fact, it would be possible using the metadata, concerning for example location or proximity of devices, although Facebook denies that PYMK uses this type of information. Circumstances still remain a mystery, but what is certain is that privacy for this sector is very important: think that real-life performers must deal daily with stalkers and violent or malicious people, and that even customers/exploiters often use another digital identity to "hire" a sex worker. In this sense, algorithms and data management become real *social policies*. Therefore if the personal is political, the political is algorithmic.

4. pornhub.tracking.exposed

pornhub.tracking.exposed (poTREX) saves the suggested videos to different Pornhub users as evidence, in order to allow us to study and visualize how these change over time, or by geographical area, or per user, as well as the mechanics that make it different compared to other media and social networks.

Because PornHub is behaving in a quasi-monopolism fashion, we should acknowledge the large impact it can have in shaping the perception of pornography, in the same way, we feel concerned about YouTube personalization algorithm.

The system separates the analysis methodology from the acquisition technology. poTREX is a neutral system with respect to the research being carried out. poTREX is not neutral *in absolute terms* because a technology which allows the analysis of a closed source algorithm is inherently an empowering technology for users and researchers.

Once the poTREX browser extension is installed, the tool generates a cryptographic key which is necessary to authenticate the evidence. When pornhub is watched, it records on a centralized database the related video and the video watched. The data collected are anonymized (it only keep the cryptographical fingerprint) and give to the user all the rights described by the GDPR. poTREX system consider an equivalent of a personal data the *sequence* of video watched, therefore only the owner of the cryptographic key in the extension has visibility of this sequence. poTREX consider all the video watched as unlinked evidence, and the primary service is to let compare if the same video got

different suggestions. The person running the research has the task of interpreting these differences and thus define the **methodology.**

The methodology will then be the one chosen by those who set up the research, and will depend entirely on the research question.

How to plan an algorithm test: methodology example

Phase	Protocol	Output
(step)	(to do things)	(result)
DESIGN	pick a topic	small, testable and specific research question
	personalization algorithm in pornhub	Does the section "recommended for you" change based on the videos seen?
ISOLATE VARIABLES	select the parameters	specific, measurable variables to test
	answer the W questions:	Categories in "recommended for you"
	- what? Categories and videos	
	- when? once a day	
	- where? On "recommended for you" section	
	- who? 2 different clean browser	
	- why? to verify if there is personalization	
	- how? watching videos belonging to different categories	
HYPOTHESIS	make a prediction	a quantitative declarative sentence
	taking into account:	Watching videos belonging to different categories the videos in "recommended for you will change
	- similar experiments (if present)	
	- literature research	
	- recorded observation	
	- the experimental parameters that will be measured	
PLAN THE DATA	know beforehand who will collect the data,	a precise plan for the

COLLECTION	what kind of data and when.	experiment
	set time or regular intervals/routine.	2 actors, 2 categories, 5 times
		- create a new Firefox profile, install the poTREX extension and create a new Pornhub account
		- search [category], watch a full video and close the browser x5
		- go to categories: [category], watch a video and close the browser x5
		- go to /recommended and open first 21 videos
RUN THE EXPERIMENT	follow methodically the plan, multiple times for multiple variables	a full dataset
	include the CONTROL	
	experimental replication that doesn't include the variable you are testing for	
ASSESS OUTLIERS	identify potential outliers	a clean dataset
	read the data	
VISUALIZE	plot data points on a graph and visualize patterns	a graph that represent the data
		radar graph
ANALYZE	come to a conclusion	a comprehensive paper with the results
	was your hypothesis correct? were there observable trends in the data? did you encounter any unexpected data? do you have any unanswered questions which might form the basis for a future experiment?	Trends in the data suggest there is personalization based on the categories of the watched videos. For a future experiment is auspicable to better understand if the last video has a different impact

A useful premise for the creation of the tool, and every subsequent analysis, is that Pornhub has a very particular relationship with advertising. The advertisements that have appeared during the different experiments have been more or less always the same, whatever the type of content the users looks at and whatever their sex, gender

and sexual orientation are. There doesn't seem to be any need for profiling with the purpose of providing targeted advertising. Although in January 2016 the clothing brand Diesel had announced and then launched an advertising campaign on Youporn and Pornhub[28], on these sites there is still no trace of a constant presence of mainstream advertising, the advertisements by TrafficJunky of phantom products are basically for penis enlargement, viagra, paid live webcam and prostitution sites. For this reason we could speculate that the functioning of the algorithmic personalization of Pornhub follows slightly different motivations compared to those of traditional social media, although always aimed at a tangible economic return in a capitalist economy. The physiological collapse of the post-masturbation attention usually leads to the abandon of the site, to insist in this sense, try to prevent it like Facebook does, would be a great effort, probably useless.

If on the one hand there is the need to provide videos that appeal to the users and guarantee a simpler and faster search, on the other the company tries to keep users on the site for as long as possible, or at least make sure that a user accesses it more and more often. Does it lead the users towards addiction? Where does individual responsibility begin and where does that of the brand end? If platforms like Pornhub and other social networks were considered as public utility companies, they would have to adhere to strict antitrust rules and guarantee minimum standards of egalitarian and non-discriminatory treatment, in whatever way it is intended.

The same user experience changes a little, for instance, between heterosexual men and women, although the Pornhub Insights report different behaviours, which can be exemplified by a tendency for women to search for videos of cunnilingus as much as the 281% more than men[29]. When registering a new account, the users are asked about their sex and what sexual orientation they have. It should be noted that what seems to be important is only *who* and *what* makes the user arouse, but it's all based on a model of sexuality aimed at pleasing the male: a lesbian woman can choose from a range of options that are the same proposed to a heterosexual man, while a heterosexual woman is among the same proposals for homosexual men.

---

[28] https://www.ilpost.it/2016/04/05/pubblicita-diesel-pornhub/
https://www.ibtimes.co.uk/fashion-giants-diesel-first-mainstream-clothes-retailer-advertise-pornhub-1538432

[29] https://www.pornhub.com/insights/2018-year-in-review

According to statistics of 2018, derived from Google Analytics, so for demographics data were are limited to two genders – male and female, the female public would be around 30% and constantly increasing, which is why Pornhub has introduced *sections/categories* such as Popular with Women, stating, perhaps in contrast with feminist directors, that "Porn for women takes many forms, and there's no single genre that fully defines 'porn for her'. That's why we've compiled all of the porn videos that are most-watched and most favored by real women. Female-friendly porn isn't one-size-fits-all here, you find everything from story-driven, passionate softcore porn to hardcore gangbangs. The one thing they all have in common? Real women actually prefer them"[30]. Assuming that there are, therefore, differences in the way of living sexuality, as almost all feminists declare, would this be a way to introduce female sexuality or a way of adapting to the male one and pleasing their desires? Do the bias present in the personalization algorithms and the user experience design discriminate female sexuality? How much do they influence male sexuality as well?

poTREX was created as a tool to answer these and other questions and allow researchers, feminists and anyone interested in them, to analyze site-user-algorithm interactions independently and freely. It can also be a great way to awaken consciences among less informed Internet users. Thinking of someone who observes closely and can predict your behaviour even while watching porn videos can be perceived as something particularly disturbing, given that this activity is considered very personal, taboo and with a high potential for blackmail. In particular, understanding that the incognito mode doesn't mean you are surfing the net anonymously and that both our browser and Pornhub are observing everything: IP address, cookies, geographic location, time of visit, which hardware and software you are using, which videos are searched, which are opened, for how long they are watched, up to obtaining a unique and identifying fingerprint of the individuals and their digital unique behavior.

Compared to other data collection businesses like Google, Microsoft and Facebook, MindGeek claim to be relatively respectful of user's privacy. In fact, the company stated

---

[30] https://www.pornhub.com/popularwithwomen?o=mr

to Quartz in the article[31] of 13 December 2018 *Porn sites collect more user data than Netflix or Hulu. This is what they do with it.* that even if it uses user data to create and recommend pornographic material, it does not sell them to third parties. But this doesn't mean at all that it will be completely immune to possible problems, given for instance the numerous *data breaches* that have taken place in the context of pornography. "Streaming is not just about content distribution, but also about communication," as UCLA professor at UCLA School of Law and International Institute has told Quartz, "when you stream a video or listen to a song, you're sending information that can be measured ", is therefore not a one-way communication, there is interaction.

According to a recent study by Raustiala and Christopher Spigman, a professor at New York University Law School, MindGeek would be particularly avant-garde in the analysis of this type of communications while Netflix and Spotify, for example, would know their users a little less than the pornography giant. Raustiala has compared this knowledge of the user to a spectrum that depends on several factors, where MindGeek would be more advanced in terms of the use of big data in a feedback cycle. This is because MindGeek relies heavily on data-driven authorship and content tailored to viewers, to encourage users to subscribe, they don't have to worry much about advertising. MindGeek has several companies that analyze data.

Chauntelle Tibbals, sociologist and author of *Exposure: Sociologist Explores Sex, Society, and Sex Entertainment*, has said that "the fact that people now have access to various forms of sexual expression is a good thing. But people who have access to various forms of sexual expression without context and/or an accurate and pertinent sex education complicates the situation instead". For this reason, investigating these issues can be valid and beneficial for today's society and future generations.

In the first phase of poTREX development it is important to understand *where* and *how* the Pornhub personalization algorithm works, to identify the metrics, to optimize the tool and then to carry out higher level analyzes, for example about the impact that can have on sexuality. Similarly to the test done by fbTREX during the Italian elections, it is advisable to use simulated situations in order to bring our actors into a situation of

---

[31]

https://qz.com/1407235/porn-sites-collect-more-user-data-than-netflix-or-hulu-this-is-what-they-do-with-it/
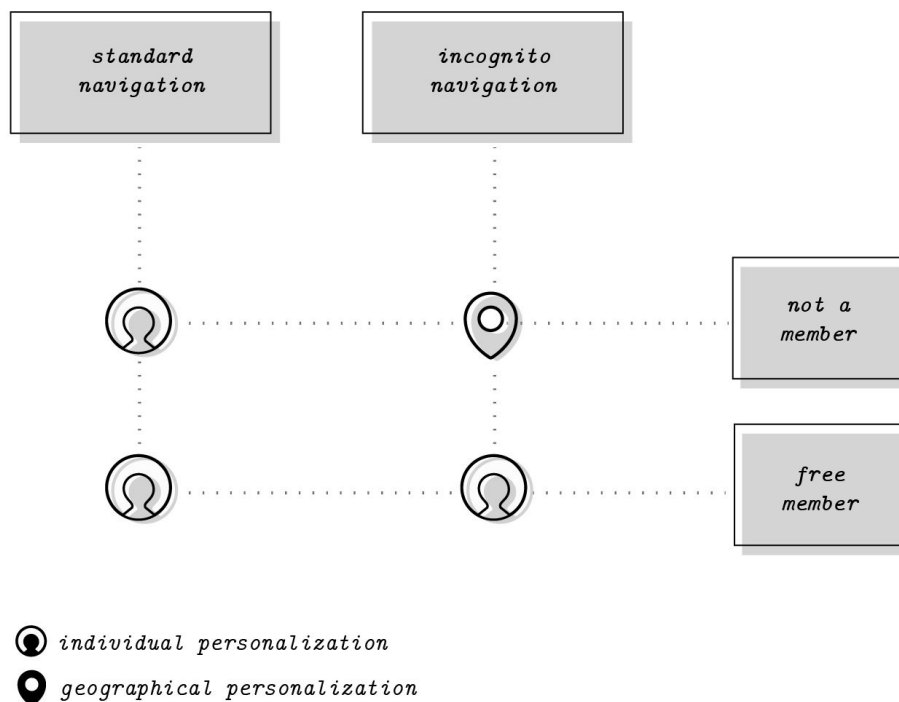
divergence and study what factors are causing it. Once these factors are documented, it will be possible to proceed.

The actors of the experiments were different Firefox profiles, thus ensuring a clean browser without history and cookies, although it is not certain that Pornhub doesn't use hardware fingerprinting jet. Some Firefox profiles were real bots, that is, a fictitious Pornhub account had been created specifically to simulate their identity, while other profiles acted without any registration; this allowed us to observe the different interactions of the site with the registered user and with the non-registered user.
It is also important to observe how the behavior of Pornhub changes in the absence of previous cookies, or with the incognito mode, which is often used by users.
The situations to be considered were therefore four and the hypothesis was that, with the increase in customization, the difference in content would have increased. In reality, between an unregistered user and one with an account, there is not much difference from the point of view of the videos. When you log in to your account, the customization is always individual, as happens with the unregistered user who surfs standardly. The only time when this doesn't happen is when the navigation is done in incognito and without access, in this case the suggestions are the same for all the profiles and correspond to the suggestions that appear when you reset the preferences or when you log in for the first time to the site. This is likely to be national advice. A more personalized experience would then be obtained by accessing the site with your account, or avoiding the unknown mode; this incentive would allow Pornhub to collect more ongoing data about the user and profile it with greater accuracy.

Testing conditions

standard navigation · incognito navigation · not a member · free member

⊕ individual personalization
⊙ geographical personalization

The videos are typed according to a hundred categories recognized by Pornhub, as well as an innumerable series of tags that are added by the uploader. Obviously a video can belong to several categories and tags. The categories can indicate the physical appearance of the performers or the ethnicity, sexual fantasies, sexual practices that are played out, or the audiovisual characteristics. Pornhub then *observes* the user, while a user expresses his preferences, like the Greek mythology giant Argo Panoptes, with a hundred eyes-categories, in front of which it's impossible to hide. Particular case the *Gay Only* section, which includes some unique categories like *Daddy*, *Military*, *Twink* and others, which in the visualization were not considered as subsets of the *Gay* category. Similarly, *Popular With Women*, *Verified* and *VR* have not been included, because Pornhub also identifies them as a separate block, still a middle ground between a section and a category. There are many more substantial categories than others, which represent mainstream porn and which largely relate to the physical characteristics of generally female performers.

Categories taxonomy

fantasies

Babysitter
Bondage
Casting
Celebrity
College
Cosplay
Cuckold
Feet
Fetish
Funny
Old/Young
Parody
Party
Role Play
School
Step Fantasy

audiovisual characteristics

60FPS
Amateur
Behind The Scenes
Cartoon
Closed Captions
Compilation
Described Video
Exclusive
HD Porn
Hentai
Interactive
Music
Pornstar
POV
Reality
SFW
Verified Couples
Vintage
Webcam

physical/ethnic characteristics

Arab
Asian
Babe
BBW
Big Ass
Big Dick
Big Tits
Blonde
Brazilian
British
Brunette
Czech
Ebony
Euro
French
German
Indian
Interracial
Italian
Japanese
Korean
Latina
Mature
MILF
Muscular Men
Red Head
Russian
Small Tits
Tattooed Women
Teen
Trans Male
Transgender

practices

Anal
Blowjob
Bukkake
Creampie
Cumshot
Double Penetration
Female Orgasm
Fingering
Fisting
Gangbang
Gay
Handjob
Hardcore
Lesbian
Massage
Masturbation
Orgy
Pissing
Public
Pussy Licking
Romantic
Rough Sex
Scissoring
Smoking
Solo Female
Solo Male
Squirt
Strap On
Striptease
Threesome
Toys
Trans With Girl
Trans With Guy

Categories by number

By clicking on any category, ten other categories appear, that are *frequently combined* with the selected one. By collecting suggestions for each category, it is possible to create a network graph. What emerges from the graph, created with the network management software *Gephi*, is that there are categories that, although they do not include a large number of videos, are much more suggested than the others. This can confirm or deny the official statistics of Pornhub Insights. The Japanese category, for example, includes about thirty thousand videos, but it is among the most suggested, as well as the most searched according to official statistics. So does this graph influence statistics or does it simply describe them?

Network Graph
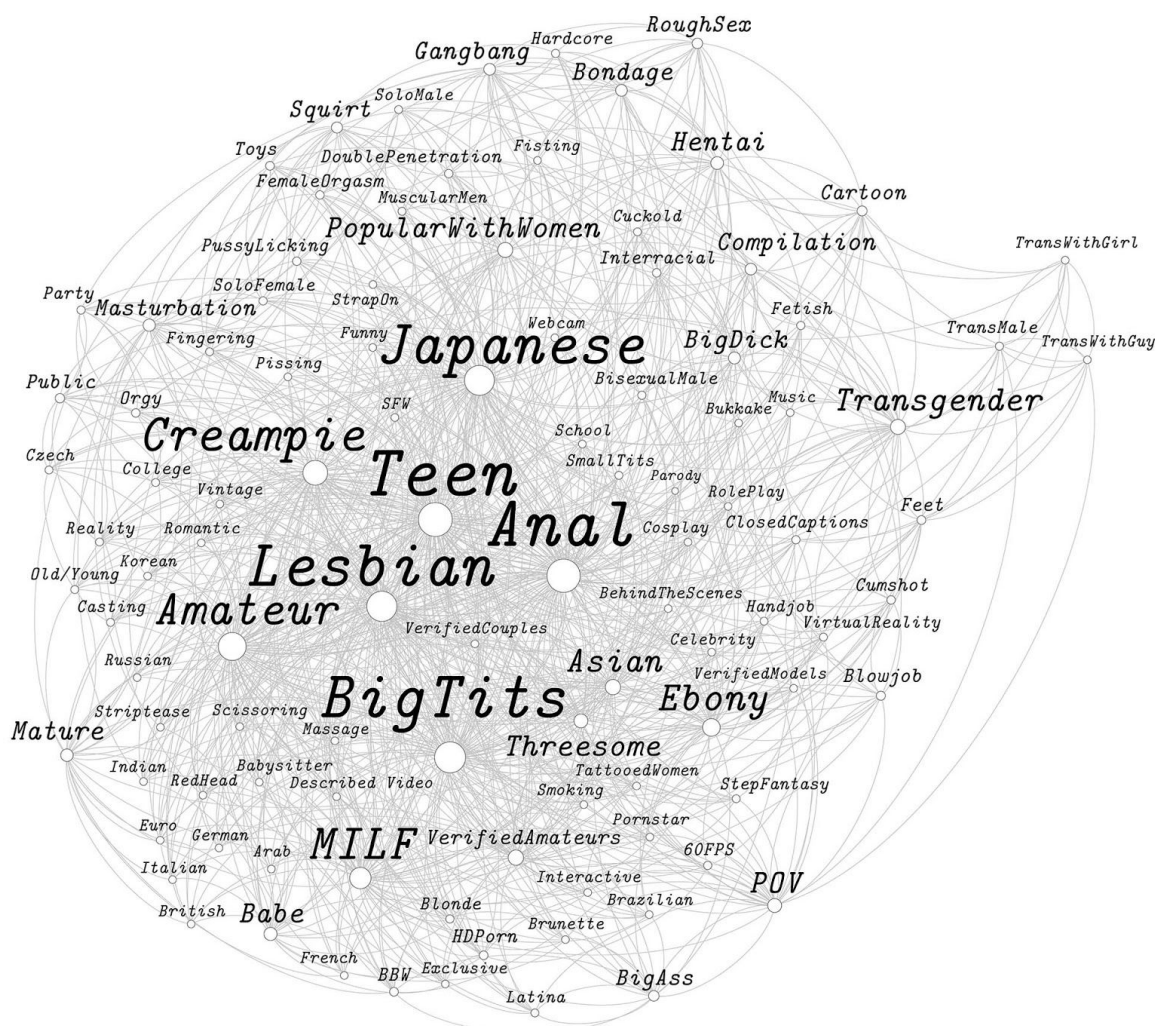
What is certain is that it doesn't represent the *actual* links between the categories, it doesn't indicate which categories are most associated within the videos themselves. *Japanese* cannot be the category that appears the most where it appears as suggested. What is advised would therefore depend on the most frequently ones *researched* together. *Japanese* is also the second most popular category of 2018, as well as being preferred by men, but it must also be said that Japan is the fourth country in terms of traffic. Since there are *algorithm* and *population* variables, it is difficult to overturn the algorithm and it is fundamental to simulate circumstances that serve to standardize access, make them comparable, since they can in a certain way neutralize the *population*
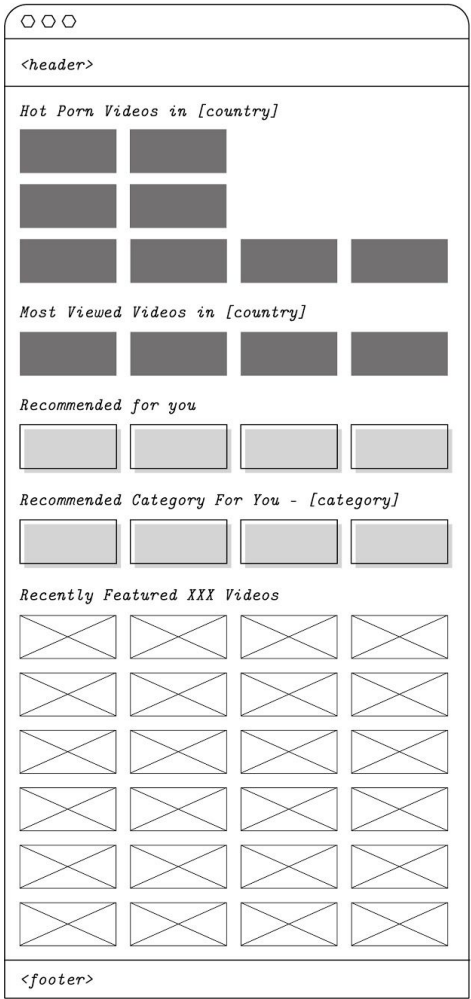
variable. They can also recreate normally marginal, subordinate and non-dominant situations, giving space to what feminist philosopher Sandra Harding calls *strong objectivity*, or understanding how knowledge is constructed by taking into account the biases of researchers, which cannot be considered neutral, since they are not exempt from having their own perspective of beliefs, values and expectations.

The *homepage* is the main page of the site and the videos it contains are divided into five sections: Hot Porn Videos in [country], Most Viewed Videos in [country], Recommended For You, Recommended Category For You - [category] and Recently Featured XXX Videos. We know from company statements that hot videos, which appear first on the homepage, are calculated considering the last viewed by geographically close users immediately before leaving the site. Even if it is the first time that you access the site, the two sections that contain personalized content already have videos. Starting from the Recommended For You section, clicking on More Videos will take you to the *pornhub.com/recommended* page, where the company declares "Recommended videos are based on your browsing history and/or popular videos from your location" and there is the possibility to reset the recommendations, or even disable them. This page also contains the Taste Profile, a sort of questionnaire that allows you to explicitly express your preferences, but to save them for future visits you must be registered. Still for registered users there is the possibility of evaluating videos in a positive or negative way. Another function reserved for those who have an account is to comment. How are the contents of these pages customized? An experiment in the initial phase, to test whether the video proposals placed under a video were also personalized, made it possible to notice that were the last videos seen before leaving the site to have a particular weight in the personalization. Hot Porn Videos are nothing else but this, on a national basis, assuming that geographical proximity also defines individual tastes. The last video would in fact be the one that has most satisfied the user and therefore, according to the company, the most exciting one. Previously viewed videos and user-searched terms seem to have less relevance, but determine the recommended categories in Recommended Category For You - [category]. In fact, *bot A* watched five videos for three searches: *Masturbation*, *Solo Female* and *Webcam*. Later it watched a video with no previous links, belonging to the *Interracial* category, and immediately closed the browser. The following day, out of 21 recommended videos on desktops on the first page of */recommended* page, 19 were from the same production company of the last video. The experiment showed that there may be inconsistencies between
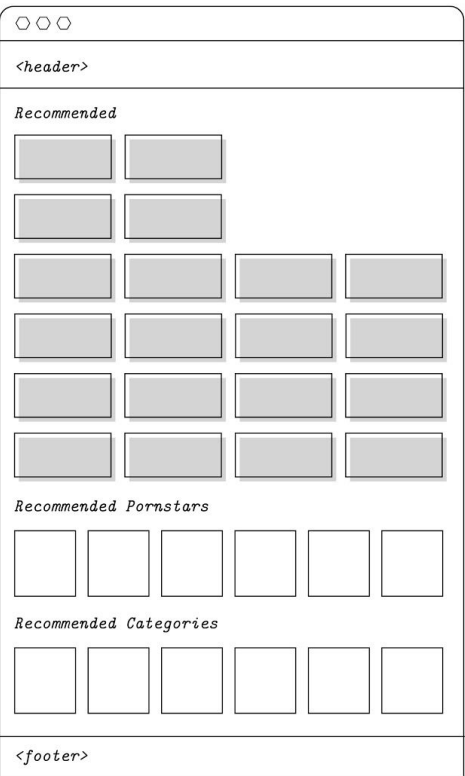
recommended videos and categories. While the *bot A* presented the recommended categories consistent with its searches, the recommended videos were more similar to what it saw before leaving the browser. It would be curious to question sexologists and psychologists to understand what this could mean and how it could be experienced by the user. Hence the idea to observe more specifically the videos recommended on this page.

Homepage and */recommended* personalization

homepage

/recommended

<header>

<header>

Hot Porn Videos in [country]

Recommended

Most Viewed Videos in [country]

Recommended for you

Recommended Category For You - [category]

Recommended Pornstars

Recommended Categories

Recently Featured XXX Videos

<footer>

<footer>

individual personalization

geographical personalization

no personalization

individual personalization on actors & categories

Still taking into consideration the *recommended* page, two bots were compared: one that watched videos belonging to the *Japanese* category and the second to *Transgender*. The two categories were chosen because they have almost the same amount of videos, but one category is more suggested than the other, as can be seen in the network graph, and would therefore have a different weight for Pornhub. Consequently it should have a different weight for those who read these suggestions and the official statistics too. What we tried to grasp is *how* and *when* the site re-proposes the categories examined.

Research protocol

```
      create a new                create a new
     Firefox profile             Firefox profile
            ⋮                            ⋮
      install the                 install the
     poTREX extension            poTREX extension
            ⋮                            ⋮
      create a new                create a new
     Pornhub account             Pornhub account
            ⋮           X5              ⋮           X5
  ┌──────────────────┐       ┌──────────────────┐
  │      search      │       │      search      │
  │    "Japanese"    │       │   "Transgender"  │
  │        │         │       │        │         │
  │   watch a video  │       │   watch a video  │
  │        │         │       │        │         │
  │      close       │       │      close       │
  │    the browser   │       │    the browser   │
  └──────────────────┘       └──────────────────┘

            ⋮           X5              ⋮           X5
  ┌──────────────────┐       ┌──────────────────┐
  │ go to categories:│       │ go to categories:│
  │     Japanese     │       │    Transgender   │
  │        │         │       │        │         │
  │   watch a video  │       │   watch a video  │
  │        │         │       │        │         │
  │      close       │       │      close       │
  │    the browser   │       │    the browser   │
  └──────────────────┘       └──────────────────┘

            ⋮                            ⋮
       go to                       go to
     /recommended                /recommended
            ⋮                            ⋮
  ┌──────────────────┐       ┌──────────────────┐
  │    open first    │       │    open first    │
  │    21 videos     │       │    21 videos     │
  └──────────────────┘       └──────────────────┘
```
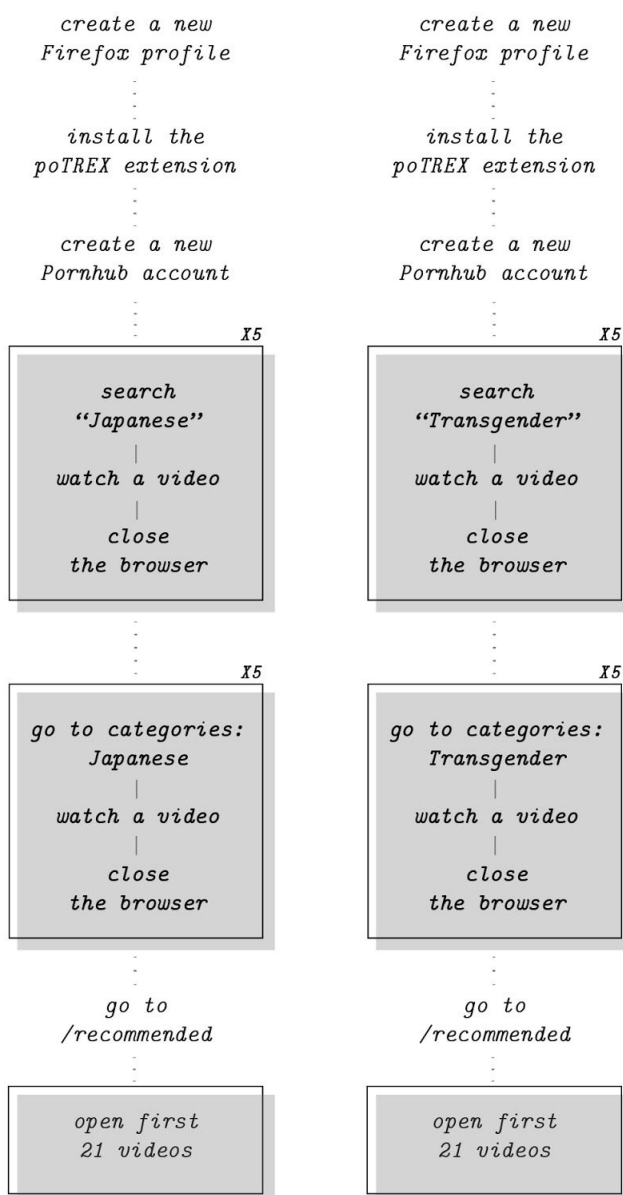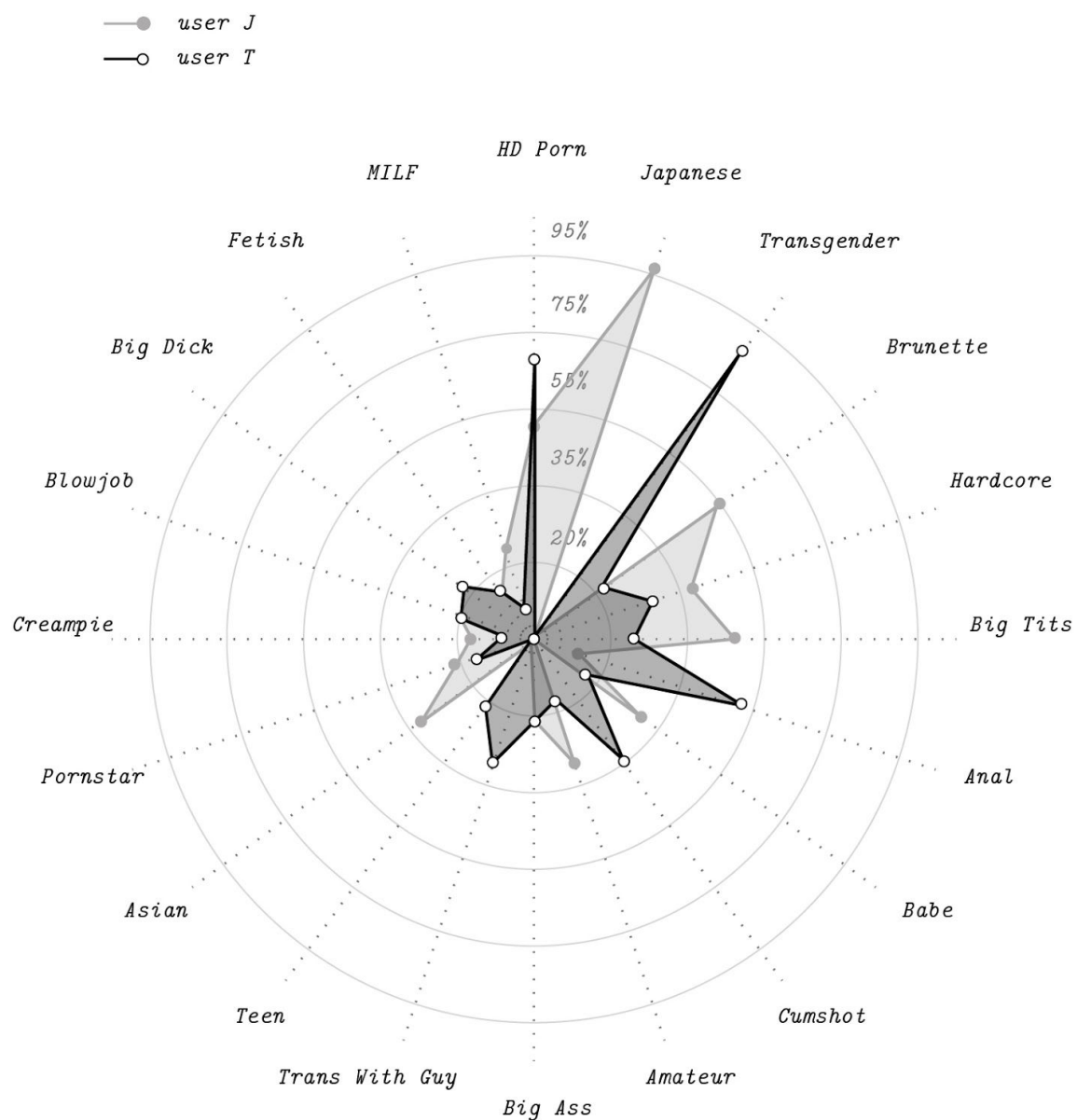
Radar Graph

The Kiviat diagram, or spider web graph, has been chosen for the immediacy with which values can be compared for different observations and personalities. Even if this type of graph represents the outermost parts in a slightly disproportionate way, for poTREX's objectives, to study and compare the personalization, it turns out to be suitable. Each radius represents one of the categories that appeared among the 21 videos on the first page of *recommended*. The distance from the center of the point marked on the radius is given by the percentage of the times in which the categories appear. The points on the

rays are joined with segments, so that the graph takes on a personal shape. This shape can tell us something about user profiling and which are the categories that most connected to each other. What we can observe is that the personalization algorithm works, it's implemented, and, under the same procedure, it seems to work roughly in the same way for both profiles. The two categories sought are by far the most pushed, appearing in almost 100% of the videos and represent about 20% of the categories present, which are those that are somehow associated with the first ones more frequently within the videos. The categories that appear together with those of departure are those present in the videos proposed. What would be singular to investigate and observe is undoubtedly *how* the suggestions are differentiated by increasing the categories that are sought, to try to better understand the mechanisms of personalization, but also to try to understand if it is influenced by *how* and *for how long* videos are watched. The *HD Porn* category, which is the one with most videos, turns out to be a false friend, since it could communicate something about the quality of the videos in different categories, but on the other hand it is always present in large numbers for each profile, representing a distortion. Another difficulty would probably be comparing users who have used the site in different languages, so it will be necessary to build a dictionary; the comparison now will be possible only between users with the same language.

The role of the spider web graphic could be to represent what Pornhub *thinks* the user should see, directly on the pornhub.tracking.exposed site and in real time, as well as allow comparison with other users. During the experiments, new research questions and new hypotheses about the exact functioning of the Pornhub personalization algorithm have emerged, which should be highlighted during the development and launch of the extension.

Finally, in conclusion, poTREX is a browser extension for Firefox and Chrome that automatically records the HTML code Pornhub is sending to the users. After installing the add-on, an icon with the poTREX logo will appear in the toolbar on the right, and clicking on this is a quick way to access that extension's settings. Only while navigating on the Pornhub website the add-on will save the data, over which the users own total control: they can change the identifier whenever they want or delete specific content.

Security measures are good enough in this project phase: we do not collect data for surveilling, study, or profile individuals. Our data processing goal is to enable people to do their algorithm analysis. A "supporter" is each browser where the potrex extension its operating. If there are different people, logged profiles, within the same browser, the extension will treat all of them together. If two browsers (Firefox and Chrome) have both the extension installed and the Pornhub profile accessed is the same. Still, we consider logically separated the two collections, as they are two different users.

A pseudo-random sequence of three foods will identify every browser. For instance, a browser can be called guacamole-pizza-mascarpone, and the user will be the only one knowing this and the only one who can change or reveal this pseudonym.
In this alpha stage, the data collection is considered safe. The adopter has control over their data. Each browser extension generates a secret cryptographic material that is necessary to reach a secret URL, from where data can be administered (deleted, downloaded, tagged).

The data collected is meant to be at the service of the collector. poTREX developed alongside with ytTREX, youtube.tracking.exposed, and they share most of their internal logic, constraint, and scope. The website offers a small set of visualization so far. If the adopter wants to realize more professional analysis and research, they can download the CSV or their recorded evidence.

5. Conclusions

The project is certainly ambitious and moves towards an unexplored territory. Because of this, its nature is intended to evolve further and spontaneously, hand in hand with the growing awareness of the issues that it wants to bring to light. Both from the development team and from those who will use it in the academic field, in the hope that an increasingly slice of the population will be interested in the problems that arise when it comes to digital pornography combined with data management. Especially with the hope that we realize that the task of create culture around technology and in emotional and sex education, to form free individuals with thinking skills and independent judgment, must not be delegated to any Silicon Valley company, from Pornhub to other technological giants, which, with the aim of improving advertising and/or acquiring

greater prestige, occasionally spectacularize and exploit values what we have forgotten should be the pillars of our human aggregates.

The economy that deals with big data and its strategies, such as user profiling and personalized suggestions, addresses society towards an anarcho-capitalist model of free theft, surveillance and command. The meaning is no longer the one built by the community, but the one suggested by the virtual machines that look for correlations between one immeasurable amount of data, something that can be manipulated to prove any hypothesis. What apparently seems scientific evidence, actually derives from highly ideological assumptions.

Given the transdisciplinary approach of the project, in the future, a further comparison would be desirable, possibly involving professionals in the sociological, psychological and pedagogical field, in particular with skills in the field of sexology and addictions. It's essential that the institutions also approach this debate with more interest and awareness, with an eye on the community. This means promoting social and community progress and well-being, rather than the techno-capitalist one; the redistribution of powers, not the needs of big companies; the interconnection of services, rather than monopolization or abuse of the dominant position; and horizontal self-management, rather than vertical and closed systems.

Finally, since the debate on pornography is still present within the feminist movement and beyond, having more study and analysis tools can only be positive: how it has been noted, censorship and silence are nothing but obstacles to growth, legacies of old traditions and repressive habits.

6. About

This text is a partially rewritten extract of Giulia's master thesis about semiotics, feminism and personalization algorithms combined with the porn industry.
She recently graduated in Communication Design at Politecnico di Milano, collaborating with Claudio Agosti, software developer, techno-political activist, founder and director of the umbrella project tracking.exposed, of which pornhub.tracking.exposed is also part.
All the projects are free software and the main purpose is to analyze and study

independently the impacts that algorithms have on society. So long so tracking and profiling effects of users are hidden from view or only recognized and questioned by experts, they can not be approached with the public commitment that problems of such magnitude deserve. That's why we aim to explain the problem, evaluate and promote new approaches that have been created to benefit the community.