



THE UNIVERSITY OF CHICAGO

LIVING IN THE HUB: A PLATFORM STUDY OF DESIRE SEMANTICS

By

Loizos Bitsikokos

May 2024

A paper submitted in partial fulfillment of the requirements for
the Master of Arts degree in the Master of Arts in
Computational Social Science

Supported through the Onassis Foundation Scholarship 2023-2024

Faculty Advisor: Prof. Michael Rossi

Preceptor: Dr. Pedro Alberto Arroyo

Πούστηδες και παλικάρια
γίναμε μαλλιά κουβάρια.

Greek Folklore saying

Acknowledgements

This project is now complete (or at least as complete as a social science MA thesis written by a “physicist” could be). Yet, it would not have been complete without the support of numerous individuals to whom I’m profoundly grateful.

First and foremost, I owe my gratitude to my first reader, Prof. Michael Rossi for his support and invaluable expertise in Science and Technology Studies and Semiotics, which composed the central part of my work.

Equally, I am indebted to Dr. Pedro Alberto Arroyo, also known as Peđro, my second reader, whose contributions transcend academic guidance. Apart from providing his Computational Social Science expertise and passion for Word Embeddings and Semiotics to the project, he added intellectually stimulating conversations and humor to my endeavors (often way beyond his paycheck). To that, I would add his unparalleled support as my Preceptor from my first day in MACSS, which made my time at the University of Chicago a bit less cumbersome.

My sincere appreciation also goes to the Onassis Foundation, whose financial support during my second year in MACSS enabled me to pursue and present this work in its totality.

I am also profoundly thankful to Prof. Rafe Stolzenberg, who mentored me and supported me greatly during my transition to the social sciences. His open-mindedness, clarity of thought, and welcoming openness of his office to my ideas have been tremendously supportive in my academic journey.

Joseph Helbing also deserves a special mention for his readiness to delve into data science ideas at a moment’s notice and his constant eagerness to discuss my work. Kaya Borlase, too, for her contributions to our initial Large-scale computing project, and Prof. John Clindaniel, whose encouragement to transform this project into a thesis was critical, being the first to see its potential while also providing me with my first references on Taboos and Stigma. Out of the people in MACSS, I also thank Dr. Sanja Miklin and Roberto Rondo Garces for their contributions to the ideation process, Prof. Jean Clipperton for helping

workshop the study's graphs, and an anonymous friend whose subject matter expertise in pornography was invaluable.

Among my friends, I wish to thank my dear friend and computing expert Gerasimos Christodoulou for, rather unknowingly, sponsoring the VPN service used in the experimental studies and for, rather knowingly, offering his whole-hearted moral support. An enormous thanks also goes to Konstantinos Panagiotopoulos for the camaraderie, the late-night calls, those hot Athenian summer days spent researching (including eating, drinking, and joking; all these count as research anyway), and, of course, his vision of the plane tree (colloquially known among few as *The Platanus*).

Lastly, but most importantly, my most profound appreciation goes to my parents, Kostas Bitsikokos and Niki Mavrovounioti for all their love, moral and financial support. Without them, I would not have been brought to this life, but most importantly, I wouldn't have crossed the Atlantic. To those curious inquiries about how they feel that their son researches online pornography, I give their unwavering answer: "Proud!".

To all who are cast in light or shadow and the remaining few who remain "pessimists of the intellect, but optimists of the will", please accept my sincere gratitude.

Abstract

Since the late 2000s, the world has experienced an unprecedented platformization of cultural production. Cultural content is created, disseminated, and consumed on online platforms, leading to a state wherein algorithmic systems and coding infrastructure primarily mediate interaction with audiovisual media. The question this thesis addresses is how we can use existing algorithmic and software systems within platforms to study sexual desire and sexuality or, at the very least, the lust of their users. To do so, I assess the social norms unfolding within the content categorization and recommendation systems of Pornhub, a popular pornographic video online platform. My methodological approach is multi-modal since observational data (collected and found) are combined with interpretation and digital experiments. Conventional content analysis methods, such as term frequencies or semantic networks, are limited in their ability to capture the nuances of video category semantics. To work around this, I use a custom word embedding model treating categories as the model's "vocabulary" and mapping them to a 100-dimensional semantic space in which similar words appear in proximity. Results showcase a promising and detailed semantic understanding of category meaning that can be further applied in studies of algorithmic bias of the platform's video recommendations.

Keywords: Platform Studies; Porn Studies; Science and Technology Studies; Algorithmic Bias; Word Embeddings; Content Analysis; Computational Semiotics; Computational Social Science

Contents

1	Introduction	1
2	Literature Review	3
2.1	“(Insert X)” Studies	3
2.2	Understanding Online Pornography in a broad context	7
2.2.1	Pornography and the Digital	7
2.2.2	Categorization in Online Pornography	9
2.2.3	Semiotics of Online Pornography	11
2.3	Algorithmic Systems, Society and Identity	16
2.3.1	Algorithms as the level of analysis	16
2.3.2	Algorithms and Society	17
2.3.3	The Self in the Digital Machine	18
2.4	Theoretical Synthesis	27
3	Data and Methods	28
3.1	Data	28
3.2	Methods	29
4	Results	34
4.1	Content Analysis	34
4.1.1	Term frequency analysis	34
4.1.2	Category trends	35
4.1.3	Semantic Network	36
4.2	Word embeddings	39
4.3	Online experiments	43
5	Discussion	45
5.1	Methodological notes	45

5.2	Category meaning space	46
5.3	Experimentation	50
6	Conclusions and Future Trajectories	52
	Appendix A Appendix	54

1 Introduction

Online platforms by claim control cultural content, from its production to distribution (Poell, 2020), and their dominance over social connections has been acknowledged and widely studied in social science research (Salganik, 2019). Software systems, increasingly mediate our connections to the world. This text, for example, was written using AI-backed spellcheckers while listening to music recommended by Spotify, watching talks on Youtube¹, scrolling on social media, and chatting on multiple messaging apps. There’s an online platform for whatever our heart desires, be it new music to listen to, films to watch, staying informed, someone to talk to, or good old carnal lust. The words of Andrew Thomson, a Components² researcher, well capture the moment: “We all live in Pornhub. We live in these various modalities that are Pornhub-like or Spotify-like” (Thompson, 2019). We all live, whether we like it or not, in this vast digital landscape of online platforms. To paraphrase Geertz’s famous definition (Geertz, 2017, p. 5): culture on the Web must be seen as the webs of meaning that platforms have spun for us .

There are multiple fruitful ways to think about platforms. Platforms can be seen as Actor-Network sociotechnical assemblages of: (i) software engineers, algorithmic systems, servers, data-bases, cloud resources, personal computers, smartphones, cameras, audiovisual content, textual elements, users, and consumers, and (ii) a complex relational web of materialities and humans much like what Glaser et al. (2021) describe as organizations enhanced by digital technology. Thus, multiple actors play a role in shaping sociality within them. Yet another way would be to follow Foote (2022) and frame online communities, and by extension platforms, as organizational systems. This “System Engineering” perspective arises primarily in organizational literature and provides an analytical lens for studying organizations as superstructures emerging from interactions of interconnected parts and their environment (Scott, 2002).

¹Interestingly enough, at this exact time, I’m listening to a 2017 discussion between Dimitris Poulikakos, Periklis Korovessis, Teos Romvos, and Antonis Antonakos on May 68, where they mentioned that their discussion might be conducted freely as it is not considered dangerous by the status quo. My hopes for this thesis are (dis)similar.

²Please see <https://components.one/>

On the other hand, one could focus on the institutionalization of online platforms following Nieborg and Poell (2018) and Nieborg et al. (2020), according to whom platforms must be seen as vital infrastructure for sociality that, at the same time, regulates and controls, curates, disseminates, and monetizes social activity.

Interestingly, often-contradictory theoretical frameworks used to study online platforms converge to the extent that each argues that they can shape individual and, by extension, social identity. As there is no limit to platform activities, they can shape identity at large, from desires to cultural and political identities. In that sense, it would be helpful to incorporate a dual distinction between the top-down and bottom-up view of platforms: on the one hand, platforms enforce cultures (users have control at best around the margins), and on the other hand, users form communication networks and communities within them. The extent to which users interacting within platforms can challenge institutionalized constraints is still an open question, the answer to which most probably depends upon platform-specific dynamics. Nonetheless, I must note that platforms often extend far beyond their superficial facade of nominal intention; they not only host content (be it music, videos, text, etc.), but also include advertisement networks and data collection mechanisms for user activity monetization (Poell, 2020).

Pornography remains at the focal point of this online space, a billion-dollar industry³ that receives comparable (and in some cases higher) levels of traffic and user engagement compared to conventional cultural online production/consumption (Wright et al., 2023). For example, Xvideos, a well-known adult video website, receives 2.9 billion visits in a given month, and Pornhub, another popular porn video website, 2.6 billion (Wright et al., 2023). To compare, conventional video consumption giants TikTok and Netflix receive 2 billion and 1.4 billion visits, respectively (Wright et al., 2023).

In this study, I investigate how we can use the platform and algorithmic infrastructure of Pornhub to study human sexuality as expressed in the digital space of online pornography.

³The exact number of the global revenue of the pornographic industry is hard to define as contradicting numbers come up over the years (see for example Center, 2023; “How Big Is Porn?”, 2001; Naughton, 2018; “Things Are Looking Up in America’s Porn Industry”, 2015).

Thus, the study deals both with the cultural and algorithmic aspects of the platform. As we will see in what follows, the distinction between the two can be heavily intertwined. I will assess how different social norms unfold within its content categorization and recommendation systems. The theoretical framework used attains analytical plurality, incorporating both a top-down as well as a bottom-up perspective of the platform while borrowing concepts from Science and Technology, Media, Software, Porn, and Platform Studies. In what follows, I conduct a semantic analysis of the video meaning space. To do so, I use video metadata and a small-scale digital experiment assessing how users can interact with platform infrastructure that helps to evaluate algorithmic bias. This work thus investigates (1) whether the platform’s categorization of content can be meaningfully interpreted based on video metadata, (2) the bias (if any) in algorithmic recommendations of different user types, and (3) whether users can escape institutional biases (if any) regarding content and reshape their sexual identity. Overall, the study involves assessing the kinds of cultural patterns arising in the platform’s content and its algorithmic recommendations.

Answering these questions has profound sociopolitical consequences as they explore the extent to which pornographic platforms like Pornhub enforce social norms and what happens when individuals proactively challenge them, thus, connecting the topics to broader sociological debates around autonomy, structure, and agency.

2 Literature Review

2.1 “(Insert X)” Studies

Contemporary socioscientific inquiry is often divided into subfields that bridge the boundaries of individual disciplines; I am, of course, referring to the various “X Studies”: fields such as Media Studies, Science and Technology Studies, Software Studies, Platform Studies, and the very focus of this project, Porn Studies. The study of online pornography inevitably borrows from multiple of these established or emerging fields of inquiry. In the following, I briefly provide the background on the respective “Studies” fields that compose my work’s

core theoretical premises.

Software Studies Software Studies is the critical research field as it has software and digital objects, or more accurately, the “stuff” of software, at its focus (Fuller, 2008). The field argues that software must be treated as a cultural object and, by extension, should be studied in terms of its politics, social aspects, ideology, and aesthetics using methods from the broad field of cultural studies. Software is also treated broadly, covering anything from algorithms, computing aesthetics, programming subcultures, and politics. Software Studies scholars stress the need to use the discipline to see through software and power structures within it (Mackenzie, 2013), while acknowledging the agency attributed to software in shaping cultural elements (Mackenzie, 2013). An intellectual space for exploring Software Studies topics is the homonymous transdisciplinary MIT book series, which generally treats how software, as a product of Culture, shapes society, culture, and identity (MIT-Press, 2024).

Science and Technology Studies Science and Technology Studies (STS) is an interdisciplinary field that analyzes scientific knowledge and technology in terms of their social and cultural implications and context. The field arises from Kuhn’s influential work in the History of Science, notably his book titled *The Structure of Scientific Revolutions* published in 1962. In Kuhn, 2012, we first see ⁴ relativist thoughts on knowledge and ideas about its social construction and a view of science as a collective culture. However, later, Kuhn views his work less under the relativist lens of some of his supporters (Bird, 2022).

Out of this line of thought on relativist epistemology, most notably Kuhn’s *Structure* and

⁴More accurately, traces of this line of thought can also be found much earlier in Ludwik Fleck’s *The Genesis and Development of a Scientific Fact*, published originally in German in 1935. Fleck (1981) defines the *thought collective* as a community of people interacting and exchanging ideas that also share a particular *thought style*, a stock of knowledge and culture. Fleck’s concepts of the thought collective and thought style are similar to Kuhn’s ideas about a *scientific community* and a *paradigm* (a set of ideas, methods, thoughts, and assumptions shared by the community). However, as Mößner (2011) highlights, despite having similarities, the two are pretty distinct. Although interesting, the arguments and historical background offered by Mößner (2011) exceed the purpose of the discussion). Nevertheless, despite Fleck’s contributions being widely acknowledged nowadays, his work received attention from the STS community relatively late, with the first English translation of *Genesis* released in 1979.

the work of epistemological anarchist philosopher Paul Feyerabend (see, for example, Feyerabend, 1975 and Feyerabend, 1987), arises the so-called *Strong Program* in the sociology of science (Bloor, 1991).

Bloor's seminal work, *Knowledge and Social Imagery* (1991) argues against the "weak" sociology of science that only targeted its sociological lens towards "unsuccessful" or marginally scientific endeavors. The Strong Program's main components include (1) **causality**: the examination of conditions that provide science with claims of knowledge, (2) **impartiality**: the need to examine both successful and unsuccessful theories, (3) **symmetry**: sociological explanations, which should explain equally true and false beliefs, (4) **reflexivity**: the method should apply to sociology itself.

The social construction of knowledge was widely accepted by the 1970s. However, around the 1980s and 1990s, the "social turn" to technology happened Woolgar, 1991. The term refers to an expansion of relativist claims and the approach of the Strong Program to technological artifacts (Woolgar, 1991). Pinch and Bijker, 1984 in their seminal work advocated for the Social Construction of Technology (SCOT). Its ideas are a push-back to technological determinism and an acknowledgment that human actions shape technology and that the latter cannot be separated from its social context. Determinism in social studies of technology assumes that technologies can affect society, as mere neutral tools that are not, in turn, affected by society. Within a deterministic view, technological artifacts are deterministic since their use is not part of any social context; one who understands technology can also define its social effects. In contrast, social construction posits that humans create technology, which in turn carries culture, ideology, and its own goals; technology affects society but is also affected by it in return.

Platform Studies Platform Studies is a branch of Media Studies that borrows concepts theory and methods from multiple disciplines (Software Studies, Organizational Analysis, continental approaches to Business, Sociology, Anthropology, Political Science, etc.). The field treats online platforms as its object of analysis and focuses on the contemporary plat-

formization of cultural production, i.e., the fact that digital platforms produce, distribute, monetize, and regulate cultural content (Poell, 2020). The field highlights that corporations are hiding behind platforms that are, in turn, embedded in broader organizational ecosystems. As a consequence, platforms assume a commercial character that includes user profiling, targeted advertisement, and Big Data collection (Poell, 2020).

Bogost and Montfort, 2009, editors of the MIT book series with the same name, provide some context on the field’s goals, aspirations, and methodology that resembles an intellectual manifesto. Most importantly, the discipline stands against “hard” versions of technological determinism. In that sense, Platform Studies borrows the concept of the Social Construction of Technology (SCOT) from the Science and Technology Studies (STS) field.

Even so, Platform Studies critically sees some aspects of the “Strong Program”. As Bogost and Montfort, 2009 argue, constructivist approaches to technology usually focus on the technology’s creation but not its use and operation within society. Platform Studies, although acknowledging the importance of constructivism in a technology’s conception, argues that equal focus should be given to its material construction and use; in other words, it advocates for a so-called “soft” version of technological determinism.

In general, the field focuses on the relationships between hardware, software, the design of computing systems, and the creative work produced within them (Bogost and Montfort, 2009). Adhering to this “soft” version of determinism, the platform studies paradigm holds that any social process within the platform is supported or limited by the platform infrastructure (front and back end design, user interface, database systems, functionalities, features, and so on).

Porn Studies The final “Studies” field that the work presented here is in direct conversation with is, naturally, *Porn Studies*. In her seminal work on the field, Williams (2004) discusses many of its premises and goals. Diverging from earlier feminist discussions around sexual politics or pornography’s legitimate right to exist (see, for example, the relevant discussion offered by Duggan and Hunter, 1995 and Segal, 1998), Porn Stud-

ies focuses on analyzing pornography as a cultural genre (Williams, 2004). Broadly, Porn Studies acknowledges the prevalence of sexually explicit or implicit⁵ materials in modern and contemporary societies. Williams (2004) highlights that the research program goes beyond treating pornography as a visual limit to film and media, which was generally the case within the work of culture theorists like Zizek (1992) or Barthes (1981).

Porn Studies treats Pornography as an inextricably linked part of contemporary cultural production. Porn Studies scholars further stress that pornography surpasses Hollywood in terms of total number of films produced per year, approximately 10,000 ~ 11,000 compared to about 400 (Keilty and Leazer, 2014; Williams, 2004) and stands as a global mass media phenomenon that doubles as a billion-dollar industry (Keilty, 2018; Keilty and Leazer, 2014; Williams, 2004). There is, thus, a need to approach it as a cultural form that is an inseparable part of modern American culture, has aesthetic qualities, and has profound social implications affecting how people perceive sexuality and desire.

2.2 Understanding Online Pornography in a broad context

2.2.1 Pornography and the Digital

Keilty (2018) argues that the study of pornographic platforms is a fruitful scholarly endeavor that can unmask the contemporary media landscape in capitalism. In the digital era, platforms can massively construct and regulate sexuality and desire through algorithmic categorization (Keilty, 2018) while they document (through the collection of user data) formerly (and potentially currently) stigmatizing and taboo sexual practices, such as masturbation; an action that used to be socially classified as private, guilty, shameful or even disgusting (Keilty and Leazer, 2014).

The platform's capacity to construct and regulate desire through data should be further put into the context of Zuboff's (2020) surveillance capitalism, the commodification of user data and their use by platforms to make user activity predictable and modifiable. Indeed, pornographic online platforms collect viewer data to construct, delineate, and make

⁵See, for example, the work presented by Gentile (2010) in *Sexing the Look in Popular Visual Culture*.

predictable sexual desires, with the ultimate goal being profit (in both the traditional and surveillance capitalist sense: surplus value and data). Again, note that surveillance is broader than just pornography, as Kordela (2023) argues contemporary society is characterized by the mediation of information in all kinds of production, destroying the divide between leisure and work time and even more highlighting Marx’s insights around the capital’s vampire-like blood thirst.

The dominant presence of the pornography film industry is widely acknowledged (Keilty, 2018; Keilty and Leazer, 2014; Williams, 2004); it seems that the “internet is for porn” meme that has a group of muppets sing ⁶:

The internet is really really great, *For porn/*

I’ve got a fast connection so I don’t have to wait, *For porn Huh?/*

There’s always some new site, *For porn!/*

I browse all day and night, *For porn!/*

It’s like I’m surfing at the speed of light, *For porn! Trekkie!*

is validated as pornographic online platforms have comparable and, in some cases, higher levels of traffic and user engagement than their non-pornographic cultural production counterparts (Wright et al., 2023). Wright et al. (2023) give characteristic examples: Xvideos and Pornhub, two well-known adult explicit content video websites. Xvideos is estimated to receive 2.9 billion visits in a given month, while Pornhub adds another 2.6 billion. In contrast, the non-pornographic video consumption giants TikTok and Netflix receive 2 billion and 1.4 billion, respectively (Wright et al., 2023). The two pornographic platforms also showcase lower bounce rates (percentage of users viewing only one page per visit), higher numbers of pages per visit, and higher average visit durations (around 8 minutes).

Pornhub, in particular, produces yearly data-oriented reports around statistics and patterns arising from user activity (see, for example, Pornhub, 2022a). According to these

⁶For a complete analysis of the origin of the meme, please refer to <https://knowyourmeme.com/memes/the-internet-is-for-porn>.

reports, the platform is conspicuously popular and receives traffic from around the world, from the Global North to the Global South, East to West (Pornhub, 2022a, 2023a, 2023b). In addition, according to auditing studies and the platform itself, user profiling is based on data collected in the mother’s company network of platforms⁷(Rama et al., 2023). Rama et al., 2023 methodology includes conducting auditing studies on Pornhub’s video recommendation algorithm to assess its biases based on different user profiles (Rama et al., 2023). Additionally, Rama et al., 2023 argue that Pornhub’s content, platform politics, and profiling reinforce heteronormativity.

2.2.2 Categorization in Online Pornography

The categorization of content in studies of porn has received some scholarly attention. However, the focus is either on *ad hoc* categories defined by researchers (Sun et al., 2015), user-reported sampling (Hald and Štulhofer, 2016), or studying the genre itself (Attwood, 2002; Morichetta et al., 2019). Only recently has attention been paid to which categories the industry uses (Stegeman et al., 2023; B. M. Watson, 2021). Quantitatively enhanced studies of these categorizations mainly focus on category counts and co-occurrences (Mazières et al., 2014; Stegeman et al., 2023) and less so on uncovering the meaning behind categories without solely resorting to English semiotics and social perceptions about the categories themselves.

The recent study by Stegeman et al. (2023) offers a comprehensive analysis of classification systems in webcam platforms. Stegeman et al. (2023) anchor their discussion in broader conceptualizations of how categories shape how people navigate and make sense of everyday life with insights borrowed from Hacking, Bowker, and Star (Bowker and Star, 2000; Heller and Sosna, 1986). Traces of the importance of categorization in those terms can even be traced back to Aristotle’s attempts to categorize, well... everything(!) through

⁷Pornhub is part of the porn “tube“ ecosystem, a collection of platforms resembling YouTube, the well-trafficked online platform for sharing videos, with the difference that the videos shared in them are almost exclusively of pornographic content. It is also worth noting that Pornhub (as well as YouPorn, RedTube, and Tube8) is part of Aylo (former MindGeek and Manwin). This adult entertainment company also controls content creation channels like Brazzers or Realty Kings. For an extensive list of Aylo’s products, see <https://www.aylo.com/brands/>.

classes of *Categories* (*Κατηγορίες*) (Aristotelis, 2011).

Stegeman et al., 2023 suggest that porn and webcam platforms alike result in hyper-categorized classification systems. They offer the concept of a *categorization regime*: a collection of platform-specific categories that shape the environment of sex workers and create conceptualizations of desirability (Stegeman et al., 2023). While categorization takes part in a pre-platform context of desire (for example, think of the typical dating question “What’s your type?”), categorization remains implicit, fluid, and loosely defined. Categorization regimes differ from these “analog” categorizations in that they provide a detailed, clear-cut, and rigid set of categories that do not arise from individual or collective social processes but rather from the platform’s infrastructure.

It is worth noting that cultural elements such as sexuality are generally of a dynamic and contextual nature (Agocha et al., 2014). The emergence of online pornographic platforms through content categorization might limit this cultural plasticity. Pornhub, in particular, includes predetermined options for the gender and sexual orientation of its users as well as pre-constructed categories for the content on the platform. Although gender and sexuality categories are much more diverse than conventional platforms, the extent to which the broader quantization delimits desire or whether users can rescue meaning-making social processes (e.g., shape their gender, sexual identity, desires, and fantasies) is an openly debated question.

Despite categorization being an essential part of platforms that helps users navigate content (Stegeman et al., 2023; Wong et al., 2020), within the pornographic space, it presents a unique oxymoron. Categorization can celebrate diversity in terms of gender, ethnicity, and body type while at the same time reinforcing discrimination and fetishization of groups across these spectra (Stegeman et al., 2023). Indeed, classification can potentially diversify the representation of sex in online space as suggested by both the Stegeman et al. (2023) discussion on “Bondage”, “Old/Young”, and “Fetish” as well as platform reports on the increasing popularity of “Hentai”⁸, and “Transgender” porn categories (Pornhub, 2023b).

⁸A genre of Japanese animation, depicting sexualized characters and intercourse

Even so, the classification system of pornographic platforms is simultaneously rigid, assuming a male heterosexual audience desiring white cis-gendered women (Stegeman et al., 2023) and potentially oppressive and limiting (B. M. Watson, 2021). For example, no specific category for white women exists in contrast to racio-ethnic categories like “Asian”, “Latina”, “Ebony”, “Indian”, etc. Additionally, gay content or content targeted to women is either categorized independently or placed under a single broad category (e.g., the Gay version of Pornhub or the “Popular With Women” category). This unbalance is also evident between categories signaling bisexuality. A “Bisexual-Male” category exists in Pornhub, while there is no corresponding category targeted to bisexual women. Note that Aylo’s platforms⁹ (former MindGeek) provide a baseline standardization of desire and pornography genres due to the established nature of categories and their universal adoption by the sex industry (B. M. Watson, 2021).

Stegeman et al. (2023) offer tags as an alternative and counter-act towards platform-imposed classifications. Instead of categories, hashtags compose a categorization system from the bottom-up as creators construct tags while uploading content online. According to Pornhub, tags enhance and aid search results (Pornhub, 2022b). As a result, the extent to which tags successfully bypass platform categorizations asks for further investigation¹⁰.

2.2.3 Semiotics of Online Pornography

Semiotics is, at root, the study of *signs*. It deals with questions regarding how meaning is made and how reality is constructed and mediated through *signs* (Chandler, 2022). Pornographic videos and images are conventional visual signs that, as we will see in the following, can be interpreted in a multitude of ways. The introduction of pornography into online platforms creates, however, an additional semiotic system, that of pornographic video categories. Each category can be treated as a sign symbolizing, referring to, or signifying certain kinds of videos.

⁹Aylo’s platforms include Pornhub, RedTube, YouPorn, GayTube, and Tube8 among others (B. M. Watson, 2021).

¹⁰Especially since, in the case of Pornhub, the mother company Aylo (former MindGeek), also controls part of the mainstream mass-produced pornography (“Aylo”, 2023).

The interpretation of pornography as a cultural medium and a semiotic system has shifted over time. Culturally, pornography once assumed a political character and was used for satire or political critique (Hunt, 1997; O’Toole, 1998). In modern societies of the global North, pornography revolves almost exclusively around fantasy, arousal, and desire (O’Toole, 1998). Nonetheless, as a social phenomenon, it is still entrenched with meaning around sexual practices, morality, pleasure, private liberties, and individual rights (O’Toole, 1998).

A scholarly and feminist debate colloquially called “Porn Wars” (Burke, 2023), arises out of two conflicting interpretations of pornography, with one position defending pornography and the other attacking it. The two views not only arise out of opposite interpretations of pornography but also have practical implications around its regulation.

The anti-porn view is primarily divided between religious and feminist views. On the one hand, some Christian moral arguments propose that pornography goes against nature since sexual function is supposed to align with Christian ethics such as procreation, marriage, romantic life, monogamy, etc. (O’Toole, 1998). The other side of anti-porn arguments arises out of feminist critiques. The feminist anti-porn view charges porn with sexism and misogyny and stresses elements of abuse present in its production (O’Toole, 1998). It also highlights that it objectifies women, transforming them into vessels for male pleasure (O’Toole, 1998); in what is commonly known as the *male gaze*¹¹ As O’Toole, 1998 informs us, anti-porn feminist views arise out of first-wave feminism and are left behind by progressive second-wave feminism.

Pro-porn feminism’s position is that it is impossible to establish a single meaning to pornographic images (O’Toole, 1998). More generally, there cannot be an objective characterization of cultural elements; for example, we cannot unambiguously define what is sexist (O’Toole, 1998). Indeed, the ’60s and ’70s sexual revolution viewed pornography as a form of emancipation and an opposition to middle-class conventions. Yet, subsequent readings

¹¹The term *male gaze* originates from feminist theory and is attributed to Laura Mulvey’s seminal work *Visual Pleasure and Narrative Cinema* (1975). The term refers to how women are depicted in art and media from a hetero-male perspective, being sexually objectified for male pleasure.

of pornographic materials from that era are filled with controversy.

Consider, for example, *Deep Throat*, the porn cult classic starring Linda Lovelace. Initially praised as a milestone of the sexual revolution, it has since been critiqued as vulgar and derogatory (O’Toole, 1998). Pro-porn feminism additionally views porn as political in the sense that it suggests an alternative to dominant flavors of sexuality through fantasy, self-love, and varied (and at times multi-partnered) guilt-free lust, while the consumption of online pornography comes without any kind of social consequence (O’Toole, 1998)¹².

It is also worth highlighting that porn critiques stressing that it is not realistic enough (e.g., depicting safe sex practices) miss its escapist nature. The viewer escapes reality through fantasy (O’Toole, 1998). As O’Toole, 1998 notes, pornography as a film genre is similar to musicals; it need not be realistic¹³.

In particular, the relationship between reality and pornographic depictions in online pornography emerges as critical. This perspective has been significantly influenced by the dominance of the “medium is the message” aphorism within Media Studies, a concept tracing back to the seminal work of McLuhan (1964). McLuhan calls for a study of the characteristics of media rather than merely their content. The importance of the medium is also emphasized in Platform Studies, as exemplified in the discussion of Bogost and Montfort (2009) around technological determinism and the influence on treating platforms as quasi-determinist. Correspondingly, the digital medium of online pornography inevitably dominates its interpretation. Despite cyberspace making porn widely accessible and culturally acceptable, as O’Toole (1998) informs us: pluralism is far from a “bedfellow” of porn. Expanding on limitations to the collective imaginary imposed by pornography going online, Attwood (2010) highlights how sex is turned into a consumer product. Even aspects of the porn production industry itself, once marginalized and part of counter-culture, become gentrified and business-like as is exemplified by pornography going online (Attwood,

¹²As Attwood (2010) highlights “porn online is reconfigured as a private space within a public environment”.

¹³In fact, musicals can be theorized as *hyperreal* following Baudrillard’s conception (see, for example, Macpherson, 2020). Pornography has also been recently studied through a Baudrillard lens (see for example Mokhtari and Malek Ahmadi, 2017; Sebikova, 2018; Tan, 2018).

2010). For Attwood (2010), the increased participation of communication and media technologies in everyday life blurs the line between “real” and “representational”, which is in turn exemplified through online self-representation. Our digital traces become part of our digital selves; whether they represent our real *Ego* or assume a *Simulacrum*¹⁴ nature is a question that remains open. Independent of a final consensus regarding the digital self, Attwood (2010) raises serious and valid concerns as digitality ties sex to the discourse of consumerism, aesthetics, the expression of self, as well as sex-positive and queer politics (Attwood, 2010).

From this by-no-means extensive discussion, it should be evident that pornography’s meaning is hard to pin down. As a writer (and a citizen devoted to democratic principles¹⁵), I abstain from taking a stance on debates over regulation. Nonetheless, pornography’s introduction to the Cybersphere raises questions about its assumed polysemy, i.e., the co-existence of multiple meanings in a sign that leaves it open for interpretation. Watching porn is, at times, a polysemiotic experience open to interpretation in which viewers themselves define the meaning of its visual elements.

Some might view pornography as emancipatory (Kubicek et al., 2010; Sherman, 1994; M. A. Watson and Smith, 2012). In contrast, others might focus on the objectification of the female body (McKee, 2005; Vandenbosch and van Oosten, 2017) and the fetishization of race and ethnicity (Cornyetz, 1997; Felkins, 2015; Holmes, 2016; Pires, 2024), the expectations arising out of the pornographic representation of sex (Ainsworth-Masiello and Evans, 2019; Goldsmith et al., 2017; Meehan, 2023). For others, pornography acts as a space, free from social constraints, that enables the exploration of fantasy and, by extension, provides ways for the radical imagination to flourish (Garlick, 2011). An approach that takes polysemy seriously must acknowledge the co-existence of such views (see, for example, Daskalopoulou and Zanette, 2020).

¹⁴The notion of *Simulacrum* has its roots in Plato’s thought regarding the nature of representation (the allegory of the cave), yet is introduced to cultural studies through Baudrillard’s conception of the *hyperreal*. The *Simulacrum* is a meta-sign, perhaps resembling the original yet, having lost its connections to the original reference, forms its own reality (Baudrillard, 2001).

¹⁵Note the choice of words here, a citizen rather than an individual.

However, one must not forget that since pornographic materials are now distributed primarily through algorithmically enhanced online platforms as well as gig economy platforms (e.g., OnlyFans), content is categorized, commodified, and directly controlled or mediated by capitalist corporations. Although polysemy can still be attributed to processes of online pornographic content consumption, the content's production by corporations (which users control at most around the margins, if at all), as well as the perceived surveillance of porn viewing by those same corporations, gives pornography an additional counter-liberating character and limits the viewer's capacity to participate in meaning-making procedures.

In contemporary consumerist society, especially within sectors controlled by a few large capitalist enterprises, the practices of product creation remain unseen from the end-user's point of view. Consider of a simpler and more graspable example, the smartphone. A smartphone user who uses it daily rarely thinks about the complexity of the supply chains behind it, the sourcing of its materials, the labor put into its production, or the intricate web of algorithms and data management systems embedded in it. Additionally, the dominance of a few major firms in the market narrows the alternatives available to consumers. At least, with the lack of or difficulty of resistance from the bottom-up ¹⁶, the market will effectively constrain users within a set of options curated by its organizational participants, perpetuating a cycle of dominance.

Similarly, the continuing dominance of corporations in the production, curation, and dissemination of pornographic content through platform infrastructures, along with the limited capacities of alternative pornography, can severely limit content plurality and diversity. Similar concerns about algorithmic systems imposing and limiting social life have been expressed recently with the emergence of Large Language Models (LLMs) and their potential domination over political discourse (Bajhor, 2023). Bajhor stresses that LLM technologies could form public opinions in case they dominate the digital public sphere. This is especially warning as these technologies are completely privatized (Bajhor, 2023). Bahjor sides with concerns of Bender et al. (2021) around a possible "value lock", the fix-

¹⁶As we will see in the following, abstinence, often the last resort of resistance against surveillance, is no solution (Benjamin, 2019).

ture of social values due to AI's inability to change due to its training on past data and its prediction scheme. Therefore, it is ever more crucial to understand the meaning behind novel semiotic spaces within online pornography, such as that of video categories, to fully evaluate more profound sociopolitical debates around agency, structure, and, in particular, the capacity of online porn to liberate or oppress.

2.3 Algorithmic Systems, Society and Identity

Pornography platforms themselves have significant algorithmic infrastructure, including but not limited to hosting services, streaming video technologies, live chats, banner ads, and recommendation systems (Keilty, 2018). These, primarily algorithmic in nature, systems can shape sexuality and desire, thereby reinforcing social norms, hierarchies, and power structures. It is thus crucial to turn our sociological lens to them and explore/uncover the premises behind them.

2.3.1 Algorithms as the level of analysis

The video recommendation system of pornographic websites is an “algorithm“¹⁷ that receives as input user data provided either explicitly (e.g., gender, sexuality, etc.) or subtly (e.g., country, device, and browser used).

Assessing how algorithms work is of vital importance. For example, they may pose a threat to dynamic social processes by imposing a confounded view of the world (Pasquale, 2016). Algorithms often pose as objective truth-production systems, yet they are affected by their human creators (Diakopoulos, 2014). Algorithms can exert power, and their decisions can be political and/or biased (Diakopoulos, 2014). As Kitchin (2017) puts it, algorithms are “embedded in wider socio-technical assemblages” meaning that they are not only influenced

¹⁷The term “algorithm“ first appears in the writings of the Arabian mathematician Muḥammad ibn Musa al-Khwarizmi referring to step-by-step methods for conducting algebraic relations (addition, subtraction, etc.). The term received its modern meaning during the 60s with the emergence of high-level programming languages (Miyazaki, 2012). An algorithm is a sequence of steps consisting of a computational procedure, taking values as input and producing values as output. Algorithms can also be thought of as problem-solving tools. The problem is defined as an input-output relationship, and the algorithm is a computational procedure to achieve the desired relation. (Cormen, 2009)

directly by social processes and the biases of their creators but also the 2nd level biases of technologies used in their development.

According to Kitchin (2017), algorithms include practices such as *researching concepts*, *selecting and cleaning data*, *selling products*, and *building teams*. Those practices are embedded in systems of knowledge, organizational and institutional cultures, politics, legal procedures, and communities (Kitchin, 2017)¹⁸. Kitchin (2017) also distinguishes their view from the traditional “black box” view, such as Diakopoulos’ (2014) view of algorithms. They view algorithms as embedded in a human network with “hundreds of hands reaching into them” (Kitchin, 2017). Taking everything under consideration, and as Geiger highlights, code is a material entity that cannot be separated from the conditions in which it is developed (Geiger, 2014). Algorithms “construct regimes of power and knowledge”, constrain possibilities, and model human behavior (Kushner, 2013). There is, thus, the need to understand how exactly algorithms may shape knowledge and manipulate social processes and human behavior. In the context of pornography, algorithms can shape sexual fantasies and perpetuate taboos, stigma, and heteronormativity.

2.3.2 Algorithms and Society

The debate around algorithmic systems has, at least, two strongly opposing views. As Mitrou (2023) informs us, the discussion mainly revolves around a techno-futurist utopian view oposed by a dystopian view of technology. Mitrou (2023) discusses Artificial Intelligence (AI) as both a “general purpose technology” much like steam or electricity, in the sense that multiple sectors can use it, and the technology itself holds the keys to human prosperity. On the other hand, AI can be studied as a dystopian, de-humanized “disruptive” technology that poses dangers for society if left unregulated. Mitrou (2023) stresses

¹⁸Among organizational theories emphasizing the importance of norms and culture in shaping organizations *neo-institutionalism* is prominent (see Powell and DiMaggio, 2012). The neo-institutional approach stresses how culture and norms shape actions, behaviors, and organizational structures (Bandelj and Morgan, 2015). According to Scott (2002), “Institutions are composed of cultural-cognitive, normative, and regulative elements that together with associated activities and resources provide stability and meaning to social life”. For a detailed analysis, refer to the work of Powell and DiMaggio, *The New Institutionalism in Organizational Analysis* (2012).

the complex interplay between algorithmic systems and their creators, acknowledging that humans are involved in every step of their construction, from data to technological systems. In that sense, their results are difficult to predict¹⁹.

More often than not, algorithmic systems are biased (Tassis, 2023). Tassis (2023) defines the term biased AI as the phenomenon in which cold non-human computational programs (that are also “ethically neutral”) are based on faulty algorithms, insufficient data, and as a result, develop biased results based on a certain number of characteristics, individuals or groups. The problem, according to Tassis (2023), is that machines do not contain affective attributes, a necessary component for human decision-making. As a result, algorithms cannot adapt to a social environment requiring ethical judgments. Humans deal with ethically complex situations using their instincts, consciousness, and empathy, while algorithmic systems need intelligible commands in a language they understand (Tassis, 2023).

So far, I have discussed the social nature of algorithmic systems, examining how they could be studied from a sociological perspective. I have also presented some of their possible effects on society. However, our interactions with algorithms beg for a deeper reflection on the essence of being human. In fact, algorithms are linked to identity and can serve as mirrors of our social selves. This topic is the primary focus of the following section.

2.3.3 The Self in the Digital Machine

The relationship between digital technologies and society can be greatly explicated through the seemingly disparate analytical lenses of Ruha Benjamin in *Race After Technology* (Benjamin, 2019) and Sherry Turkle in *The Second Self* (Turkle, 2005). Benjamin examines racial biases and structures of domination embedded in technological systems. Turkle focuses on the psychological impact of computers on humans and narrates an existential account of what it means to be human in the dawn of computers. Drawing from both sources, identity is seen as constructed through interactions with technological systems,

¹⁹An insight the great philosopher Edgar Morin stresses in his most recent book: “history is non-linear; we must take into account *contingency* in our calculations, the fact that objects through their interactions may have the exact opposite result from the one we initially planned” (Morin, 2022).

while technology can both at times liberate and oppress, envision and designate, and create a world that appears like a futurist dream or a dystopian nightmare. This section will be dedicated to both of their contributions in detail. By putting them in conversation we can better understand the connections between the self and the other, represented here by digital technologies.

Race After Technology Benjamin opposes a naïve view of technology that is devoid of social meaning. She suggests that algorithmic technologies may pose as rational scientific entities but enforce and reinforce all sorts of inequities, focusing primarily on racism (Benjamin, 2019, p. 2-3). Precisely because these technologies appear scientific, they can reproduce inequity in a particularly problematic way; they pose as progressive when, in fact, they are not. As Benjamin put it: “Codes, in short, operate within powerful systems of meaning that render some things visible, others invisible, and create a vast array of distortions and dangers” (Benjamin, 2019, p. 7). Benjamin thus breaks the false dichotomy between an *a priori* neutral and *a posteriori* biased technology offered by Tassis (2023); there is no such thing as a neutral algorithm.

The dangers are evident: technological systems compose an “elaborate social and technical apparatus that governs all areas of life” (Benjamin, 2019, p. 11). Overall, regarding technologically enhanced systems of authority, a social elite quantifies and atomizes humans (Benjamin, 2019), and we are eager to give up our rights for services in advance (Zuboff, 2020). As Benjamin (2019) stresses, abstinence is no solution to this parallel of Marx’s opium of the masses (Marx, 1975, p. 175); we are not all in an equal position to revolt²⁰.

Benjamin explicitly opposes technological determinism (Benjamin, 2019, p.40), the view that technology affects society without being influenced by it in return. A deterministic view overlooks the interactions between society and technology, how its uses diverge from original intentions, or how its use leads to unanticipated outcomes. The non-determinist

²⁰This is inherently an oxymoron as abstinence, a principle of inaction, is seen as revolutionary action; a revolt against social media dominance over everyday life. A more optimistic, simultaneously simplistic view stressing the need for social media abstinence is shared by philosopher and artist Jaron Lanier (Lanier, 2018).

view sees technology as containing meaning instilled by people who use and interact with it.

Early in the book, she defines the *New Jim Code*²¹: “*the employment of new technologies that reflect and reproduce existing inequities but that are promoted and perceived as more objective or progressive than the discriminatory systems of a previous era*” (Benjamin, 2019, p. 5). Benjamin discusses four dimensions of this New Jim Code: (1) **Engineered Inequity**, the explicit attempt of technologies to reinforce social hierarchies (class/gender/class), (2) **Default discrimination**, a more nuanced approach to bias, in which the technology’s creators fail to understand the social preliminaries to their work, leading to biases “by mistake” or arising out of ignorance, (3) **Coded exposure**, the exposure of individuals through technological systems which is inevitably tied to surveillance, and (4) **Technological Benevolence**, the attempts to address issues of discrimination that end up reproducing it or making it more complex often due to a limited understanding of the underlying social context.

Benjamin suggests that we reconsider our relationship with technology, how we engage with technological artifacts, and how these interactions shape our understanding of the social world. She focuses on race, which she conceptualizes as technology in its own right because it provides a means to “*sort, organize, and design a social structure*” (Benjamin, 2019, p. 91). A core part of her argument is that technologies are not neutral or innocent and can be racist (Benjamin, 2019, p. 62). Throughout Benjamin’s analysis, multiple rich examples are developed of various aspects of the New Jim Code: Netflix racializing its movie recommendations based on user profiling (Benjamin, 2019, p. 18), Googling the term “Black Girls” providing pornographic images of race fetishization (Benjamin, 2019, p. 44), AI enhanced beauty contests determining white contestants as the most beautiful (Benjamin, 2019, p. 49), a soap dispenser sensor not able to detect darker skin tones (Benjamin, 2019, p. 68), the name of a street named after Malcom X appearing as “Malcom Ten” in Google Maps (Benjamin, 2019, p. 78), predictive policing software targeting non-

²¹The concept is named after the racist system of US laws enforcing segregation and White supremacy during the 19th and 20th centuries. (Benjamin, 2019, p.91)

white communities (Benjamin, 2019, p. 82, p. 112, p. 121), a webcam face tracing software failing to follow black subjects (Benjamin, 2019, p. 108), an image detection tool classifying black people as gorillas (Benjamin, 2019, p. 110), enforcing identity digitization as a means of control for marginalized communities in the UK, India, and Kuwait or enforcing genetic ancestry on asylum seekers (Benjamin, 2019, p. 128-129, p. 132), and, finally, machine learning enhanced job hiring processes enforcing biases (Benjamin, 2019, p. 140). All the above are just some disturbing examples of what the New Jim Code entails for individuals at risk.

Nonetheless, Benjamin acknowledges that the racial coding of technology is not new to the digital space. She mentions, for instance, multiple examples of pre-digital technologies that are part of racist narratives. A characteristic example would be photography, which was at times recruited to classify visual human differences and thus aid in constructing a narrative of white supremacy. Another example discussed in detail is that of “Shirley cards”. Shirley cards provided a baseline color balance for film exposure purposes and were developed by Kodak to capture primarily white skin tones (Benjamin, 2019, p. 103-105). As Benjamin informs us, it was only until manufacturers of chocolate and wooden furniture began complaining about the depiction of their products that the photographic industry paid attention to the cards’ constraints, with finally, Fuji famously “ethnicizing” Shirley cards (even though critiques of Fuji’s white-focused ethnicization must also be noted).

The Second Self From a different point of view, thirty-five years earlier, Turkle discusses the psychological effects of human-computer interaction. She takes an interpretive approach to studying humans interacting with computers through ethnography and psychoanalysis before digital technologies became part of our lives.

Turkle’s research subjects, mainly young children and teenagers in US schools, treat computers as “somewhat alive” or, as she more accurately put it, standing “*betwixt and between*” (Turkle, 2005, p. 24). The computer emerges as not simply a tool for computation or object manipulation (kids in her studies control “sprites” and other digital graphics

through pseudo-programming languages) but also as an object of profound cultural impact. By interacting with computers, humans create cultures, and thus, the computer assumes great power and appears evocative; it creates the condition for multiple almost magical things to happen (Turkle, 2005, p. 13).

Turkle understands early enough, through her observations and discussions with school children, that regardless of what the future entails for artificial intelligence, computing machines affect how we think. She discusses children's perceptions of the animate and inanimate, conscious and unconscious, living and not living. Turkle models her research around Piaget's approach of engaging directly with children and considering their perspectives and insights.

As a result, Turkle's observations are genuinely fascinating. Most importantly, in the 1980s, schoolkids seemed to be moving away from considering movement as a criterion of life. For example, in Piaget's writing, kids justify their responses that clouds are alive because they see them move. On the contrary, kids interviewed for *The Second Self* think of animate objects as having affection and consciousness: the computer is alive because it has psychological traits, it thinks, it wants to run simpler programs, it's cranky, or it makes mistakes. As Turkle describes this paradigm shift, "motion gives way to emotion" (Turkle, 2005, p. 58).

Multiple intriguing questions arise from her discussions with children: is the computer cheating? Does it come from somewhere? Does it have parents? Is it alive? The short answer to all of these is *sort of*, as descriptions of the computer's existence greatly resemble life. The computer needs to "eat" as batteries are its food (Turkle, 2005, p. 54), or it was born in a factory (Turkle, 2005, p. 54). Some even question what it goes on to do at night (Turkle, 2005, p. 59), reminding some of Philip K. Dick's "Do androids dream of electric sheep?". In some cases, this life-like narrative is complemented by expressing either the want for the computer to be alive (Turkle, 2005, p. 58) or the need to assume their agency and give them blame (Turkle, 2005, p. 271).

Through these accounts, Turkle inevitably discusses what is "unique" about being hu-

man. On the one hand, what makes us inherently human is what makes us treat computers as alive; this makes Turkle conclude, “If we are machines, we are human” (Turkle, 2005, p. 305). Nonetheless, a new romanticized idea of the human also arises. To be human is to be unprogrammable (Turkle, 2005, p. 57) and to have an emotional life (Turkle, 2005, p. 62). Machines might be intelligent but can’t love or hate; they lack an emotional world ²².

The computer in *The Second Self* is a mirror (Turkle, 2005, p. 306) in which we see ourselves, highlighted excellently in the following exert:

The story of Narcissus is usually read as a warning against self-love. Narcissus saw his image in the water and fell to his death because of his desire to touch it, to be closer to its beauty. But there is another way to understand the story. Narcissus fell in love with what appeared to him to be another. This image of that other person fascinated him because it objectified a sense of the beauty of which he had felt only a vague inner sense.

Therefore, according to Turkle and her research subjects, the computer’s otherness reflects the human condition.

The “realist” versus the “romantic” view The two works, separated by time and focusing on entirely different aspects of human-computer interaction, have significant differences. Benjamin, on the one hand, focuses on the structural effects of technology, while Turkle discusses the very nature of being human. This is partly inevitable; much has changed since Turkle wrote *The Second Self*. Digital technologies have become part of a larger power structure of large corporations and state actors (surveillance capitalism), which can dominate sociality. In that sense, the exploratory, collective, and “bottom-up” element of interacting with computers found in Turkle seems to have faded away today.

In contrast to Benjamin’s conceptualization of technology as racist, Turkle’s observations, at times, provide a liberating view of machines. For example, she mentions that

²²Even though the boundaries appear even more blurred today with AI chatbots and relationship simulators.

students who interact with computers feel they are reclaiming some agency and that computer knowledge belongs to them instead of their teachers (Turkle, 2005, p. 156). This futurist-grassroots view of technology is shared by some of Turkle's subjects who think of life as having endless possibilities since they can reprogram and change it anytime, just like a computer (Turkle, 2005, p. 156). In Turkle's words, computers become "*instruments for decentralization, community, and personal autonomy*" (Turkle, 2005, p. 171).

On the other hand, the *New Jim Code* conceptualizes the exact opposite view on technology. One well-known example is the case of Janet Vertesi, an STS scholar, who was trying to conceal her pregnancy from online marketing tools. Vertesi's husband was flagged as conducting suspicious transactions because he bought too many gift cards to pay pregnancy expenses while not being tracked (Benjamin, 2019, p. 152). In Vertesi's story, we notice how computing systems become, in fact, instruments for centralization and governance, contrary to Turkle's hopeful assertions.

By contrasting *Race After Technology* with *The Second Self*, a difference emerges between an early romantic view and a contemporary unromantic dystopian view of human-computer interaction. According to the romantic view, this interaction assumes a new way of thinking about humans; a new way of knowing. On the other hand, the realist view stresses the institutional and structural constraints that make technology an object of power used for domination and invigorating social hierarchies. Having seen the racist effects of algorithmic systems, Turkle's suggestion that the human-computer interaction can help us focus on identity, the soul, and the human spirit seems like a void promise at best or, at worst, a cautionary tale. Benjamin essentially reverses Turkle's call for humanizing machines and stresses the problems of human quantification; in a contemporary digital society, it's not the machine that has become human; it is we that became machine-like and, by extension, malleable, at least to the eyes of those holding power.

Finally, Benjamin offers a call for counteraction to fight this new system of oppression which unavoidable lacks in Turkle. In the last chapter of the book titled *Retooling Solidarity, Reimagining Justice*, she develops an abolitionist agenda by proposing resistance to the

spectacle approach of virtual environments and entertainment media towards news and violence (Benjamin, 2019, p. 170), tech worker resistance (Benjamin, 2019, p. 184), auditing and accountability studies (Benjamin, 2019, p. 185), or designing apps that expose or critique the status quo like an app targeting white collar crime (Benjamin, 2019, p. 196).

Points of convergence Nonetheless, despite their differences, we can also find numerous points of convergence between the two works. Both Turkle and Benjamin think of technology beyond the simple instrumentalist view of “technology as a tool”, while they both explicitly reject technological determinism. Their views converge on the fact that people construct the meaning of technology; in short, the social context matters.

Both thinkers seem to stress the importance of knowing who is behind the computer, algorithm, apparatus, and so on. For Benjamin, the reason why technology designers are influential is self-evident: they can instill biases and problematic views in their creations. For Turkle as well, a computer program reflects the mind of its programmer. For example, some kids create structured programs while others take an impressionistic approach (Turkle, 2005, p. 104). Some get frustrated at mistakes and bugs, while others embrace them as part of the creative process (Turkle, 2005, p. 113).

Some of Turkle’s remarks also reflect Benjamin’s understanding of the profound impact of societal forces on technology. Turkle, for example, discusses how societal gender dynamics influence how boys and girls approach coding and how they think of the machine (Turkle, 2005, p. 108-110). In that sense, much like Benjamin’s accounts of the racist technological system, Turkle acknowledges that computing reflects our social selves.

On a stylistic but substantial note, both authors draw from pop culture themes, Hollywood movies, TV series, and fiction to consider the interactions between humans and technology. Benjamin uses a lot of speculative fiction to see how we think about ourselves and presents multiple examples from TV shows. A notable example is an episode of *Better Off Ted* in which new technology installed in an office (elevators, automated doors, and water fountains) cannot detect black employees. Benjamin additionally draws from Ma-

trix, Gattaca, Elysium, and Snowpiercer to discuss the capacity of technology to dominate. Turkle also uses multiple films in her argumentation, drawing from Star Wars, Star Trek, Tron, Barbarella, and Woody Allen movies. We, thus, see the potential for imagination and art to provide heuristics and ways to understand the social world; both Turkle and Benjamin acknowledge art's capacity to anticipate change and reflect societal dynamics.

Even though Turkle's general stance toward technology is romantic, she does foresee many moments that could be connected to the negative impacts of technology, such as the ones explicated by Benjamin. Some points of early adverse implications of computing technology in human psychology involve (i) the emergence of virtual environments through gaming cultures (Turkle, 2005, p. 66), (ii) the attention economy also arising out of gaming consoles (Turkle, 2005, p. 69), and (iii) the depersonalization of interacting with computers as part of the early hacker culture (Turkle, 2005, p. 213, 217). Similarly, the spread of algorithmic concepts within society during the early digital age enabled a reconceptualization of humans as machines to be programmed, as expressed by student coders in Turkle's interviews. This also seems to foreshadow the concepts of control, surveillance, and quantification I already discussed. An informative excerpt is offered by Turkle (Turkle, 2005)²³:

I think that I am programmed like the computer. Other kids in the school aren't as programmed as me. They have to do things, but they don't have to do them in order. My mother did my programming. And the Pope. Well, not really, the priest did it. But the Pope did his.

The fundamental questions around humans and computing machines have remained the same across thirty-five years of research and thinking: who are we (humans), what are they (computers), and what does our interaction entail for society? Benjamin and Turkle provide invaluable ways to think about our multifaceted relationship that is both partly empowering and partly constricting. By fusing their work, we can better appreciate the allegory of humans and machines facing each other in a mirror; each is indeed a mere

²³Unrelatedly to the discussion, this segment, in particular, seems to be extracted straight out of a D'Orta, 1992 book.

reflection of the other.

2.4 Theoretical Synthesis

The following section is devoted to synthesizing the foregoing discussion around platforms, algorithms, and pornography into a coherent framework useful for my study. Online platforms, including pornographic ones, control the consumption, production, and distribution of cultural content while having the capacity to shape multiple critical aspects of identity (desire, sexual, political, cultural, etc.). Many have stressed that the coding infrastructure behind platforms must be treated as a cultural object and studied in terms of its politics, social aspects, and ideology.

Following the field of Platform Studies, my theoretical lens adheres to epistemological plurality. Drawing from STS scholarship, which views technology as socially constructed, I acknowledge the importance of the social context in studies of digital technologies. Benjamin's (2019) examples of the *New Jim Code* illustrate excellently that technological biases tell us more about our social selves than the technology itself. Software technologies, in fact, may perpetuate racism much more effectively than conventional forms of domination and control (Benjamin, 2019). The same may hold for other types of biases, such as heteronormativity, that are dominant within pornographic video platforms (Rama et al., 2023).

However, pornographic and conventional platforms alike include algorithms in their infrastructure. There is an increasing need to examine algorithms sociologically and understand their biases and social premises (Diakopoulos, 2014; Kitchin, 2017)²⁴. Therefore, it would be misleading to overlook how platform infrastructure imposes limits on the social life it hosts; its design might influence how biases propagate or what can be known within them. My pluralist theoretical view is in agreement with Bogost and Montfort (2009) and pushes for a “soft” version of technological determinism.

Porn Studies scholarship has also implicitly acknowledged the quasi-deterministic role of

²⁴The critique should extend beyond auditing studies, like those proposed by Diakopoulos (2014), to include theoretical along with empirical work.

platforms in constructing and regulating sexuality and desire (Keilty, 2018). In fact, recent Porn Studies research, aligning with Platform Studies, has targeted its lens toward video categories that at times promote diversity and sometimes hinder it (Stegeman et al., 2023). These semiotic systems of content categories are not the result of collective processes but are rather imposed on users, making a “soft” deterministic approach necessarily relevant. In accordance with McLuhan’s (1964) aphorism: in online porn, too, the medium is the message. Like any other human activity going online, pornography has the potential to both liberate and oppress, leaving open the question of which, if any, prevails.

As should be evident from the discussion, software technologies are far from mere “tools”. They carry social meaning even to the point of re-conceptualizing humans as thinking machines and vice versa. Nonetheless, the transformation of online spaces into platform economies stresses the need to treat them quasi-deterministically. This conceptualization re-introduces social structure into the equation via platform design. Rather than a “tool”, a more fitting metaphor for the algorithmically-enhanced online platform is that of a “prison cell”; to escape living in all the various Pornhub-like platform modalities is, if at all possible, a cumbersome process.

3 Data and Methods

3.1 Data

Throughout this analysis, I use a combination of custom-collected and found data in order to assess the meaning of pornographic video categories on Pornhub. I further apply this approach to study elements of algorithmic bias.

First, I use a dataset of 149,431 videos and metadata, collected as part of a collaborative group project that also involved Joseph Helbing, and Kaya Borlase²⁵. Note that this large-scale data collection and processing project is set up on Amazon’s Cloud infrastructure,

²⁵The project was part of the MACSS Large-scale Computing for the Social Sciences course taught by Prof. John Clindaniel on Spring 2023. The code and results of the initial web-scrafer are publicly accessible on Github (Bitsikokos et al., 2023).

Amazon Web Services (AWS), which makes it parallel and scalable by design. The project also contains Big Data on comments and creators. However, the analysis of these additional platform data exceeds the purpose of the work presented here. This video information dataset was expanded by a web scraper using Python’s “Selenium” module ²⁶. This was used to collect video metadata from September 2023 till March 2024. The need for an updated scraper using “Selenium”, instead of the initial scraper’s implementation though “Beautiful Soup”²⁷, rests solely on the fact that the webpage’s content is dynamic.

An additional dataset of video metadata is also used. The dataset is publicly available from Components ²⁸ and contains information on 218,004 videos, such as video titles, upload dates, video categories, and comments, among others. By joining the two datasets (the custom web-scraped and Components datasets), the scope of the study’s data was increased to 321,619 distinct videos. The variables used throughout the analysis include (i) each video’s title, (ii) corresponding categories, and (iii) upload dates ²⁹

I present a few representative data points and variables in Table 1 (the complete list of video categories can be found in Table A1 of the Appendix).

Video Title	Upload Date	Video Categories
Big Tit MILFs share a Cock and its Warm Cum!	2023-06-15	Big Tits, Blowjob, Brunette, MILF, Pornstar
House Girls - Scene 5	2023-07-03	Blonde, Pornstar, Rough Sex
Big Tit MILF needed 2 Cocks to be satisfied	2023-08-21	Big Tits, MILF, Orgy, Rough Sex

Table 1: Example video dataset values. The collected dataset also contains information on the video’s URL, the URL of the creator’s profile page, the creator’s username, the data collection timestamp, the video’s approval percentage, total views, likes, dislikes, number of times it was saved as favorite or starred at the time of collection, as well as the video’s tags and type of video production (“Professional” vs “Amateur”).

3.2 Methods

Content Analysis To explore the type of content on the platform, I initially conducted, a term frequency analysis on video categories to uncover the most and least popular categories.

²⁶See <https://selenium-python.readthedocs.io/>

²⁷See <https://beautiful-soup-4.readthedocs.io/en/latest/>

²⁸See <https://components.one/datasets/metadata-from-218000-Pornhub-videos-jan-2008-dec-2018>

²⁹The data collection and analysis code can be found on GitHub in <https://github.com/lbitsiko/pornhub-categorization> (Bitsikokos, 2024).

I also conducted a descriptive analysis of content trends across time, using either keywords present in titles (e.g., the number of times the word "teen" appears on video titles grouped by year) or category counts (e.g., how many videos of the "Teen" category there are for each year). The complete list of categories can be found in Table A1.

The content analysis is further complemented by constructing a semantic network of category co-occurrences according to the following mathematical formulation.

Let $G = (V, E)$ be the semantic network of categories, where V is the set of categories with $|V| = 128$ (for a full list of the categories see Table A1 in the Appendix) and $E = \{e_i \equiv (x, y, w_i) | (x, y) \in V^2, x \neq y, w_i \in \mathbb{N}\}$ denotes that the category x co-occurs with y on w_i videos (the edge's weight). Additionally, the Louvain community detection algorithm (Blondel et al., 2008) is applied to aggregate nodes in connected groups³⁰, resulting in four communities.

Lastly, I built a word embedding model with the dataset as a corpus, video categories as the vocabulary, and videos as documents in the corpus consisting of tokens in the vocabulary. The word embedding model maps tokens present in the vocabulary to vectors in a multi-dimensional space³¹.

The Continuous Bag of Words (CBOW) method for the Word2vec model was used (Almeida and Xexéo, 2023), using the following parameters:

1. `min_count = 20`
2. `window_size = 13`
3. `vector_size = 100`

Note that the method chosen is based on the distributional hypothesis, i.e. words appearing in a similar context share meaning³². Video categories do not have syntactical or grammatical structures and, hence, seem to partly violate the model's hypothesis. Nonetheless,

³⁰The Louvain method is based on *modularity* optimization. Modularity measures the respective density of edges within and across communities. For a detailed discussion refer to Blondel et al., 2008 and De Meo et al., 2011.

³¹The vector space is usually interpreted as a space of meaning in which semantically similar words appear in proximity (for example see Mikolov et al., 2013 or Pennington et al., 2014).

³²This dates back to Firth's dictum that "you shall know a word by the company it keeps" (Firth, 1968).

as a first-order approximation, I suggest that given a sufficiently large window size, a word embedding model can capture category co-occurrences relatively well. In fact, the proposed approach of applying embedding models to tokens lacking a language structure has been evaluated successfully (Dey et al., 2017). In this study, the window size was tuned to 13 based on the distribution of the number of categories per video (Figure A1). This was chosen to let the model grasp the complete context for each category (i.e., all the categories co-occurring with each category in a particular video).

In addition, categories corresponding to less than 50 videos on the dataset are excluded, based on the distribution seen in Figure A2. The complete list of categories excluded from the model can be seen in Table A2 of the Appendix. This is necessary as the corpus of videos does not have enough occurrences of these particular categories, and hence, extracting relational meaning from them is impossible.

Since the model’s vocabulary has 128 distinct categories, I chose to use a 100-D meaning space. At the same time, results were compared across a range of dimensions, reaching up to 300, with the findings exhibiting robustness measured by the unsupervised learning task that follows ³³

Subsequently, to facilitate the interpretation of the data and further analysis, the 100-dimensional word-embedding space was reduced using Principal Component Analysis (PCA) (Maćkiewicz and Ratajczak, 1993). PCA is a statistical technique that reduces the number of variables in a dataset into a set of variables called principal components. The principal components preserve the maximum amount of information, here conceptualized as the variance of the original dataset (Greenacre et al., 2022). The number of PCA components used was determined at seven using a Scree plot³⁴. Finally, the reduced 7-dimensional vectors

³³A conventional choice for the space dimension would be 300, as demonstrated in the work of Mikolov et al. (2013) and Pennigton et al. (2014). However, I opted for 100 dimensions to maintain parsimony. This decision was informed by the size of the vocabulary (128 categories), which is quite small compared to a typical English corpus that includes thousands of words. In any case, attempts to increase the dimensions up to 300 performed equally well, additionally making the choice of 100 dimensions preferable as less computationally expensive.

³⁴A Scree plot is the visualization of the percentage of variance explained by each number of dimensions (Cattell, 1966). The optimal number of dimensions is located by identifying the line’s inflection point (Greenacre et al., 2022).

were grouped into clusters using a K-means algorithm (Ahmed et al., 2020); the number of clusters was determined to be seven using the Elbow method ³⁵. Visualizing the categories clusters is used on the $PCA - 1 - PCA - 2$ plane (which also preserves the maximum amount of variance among the principal components) to aid the interpretation of results.

Bias experimentation In the second part of the study, I use synthetic Pornhub accounts (bots) and analyze their video recommendations. Using a VPN, I accessed the website through a pre-determined location to isolate effects from location-based profiling. The location specified is a large metropolitan city on the U.S. East Coast with a diverse population. This ensures that the recommendations extracted are reflective of a varied user base and are likely to be similar to other urban environments. This standardization might limit the ability to observe regional differences, yet does not detract from the high-level goal of analyzing core differences in recommendation patterns for certain types of users.

Six bots were created in total, simulating three male and three female users of straight sexual orientation. The bots were used for logging in to Pornhub, and their video recommendations were collected (a schematic depiction of recommendations is presented in Figure 1). Two recommendations sections are collected: the “Hot videos in your Country” and “Recommended videos” sections. These specific categories were selected based on their position on the website’s user interface design and the relevance of their topics to the study. The “Hot videos” are part of the platform’s homepage, i.e. by design the first videos a user encounters, while the “Recommended videos” signify some level of user-specific customization (making the section relevant for this particular study).

To simulate user activity and assess whether it affects recommendation patterns, each bot was made to watch the first video of a category that signifies non-normative content for its gender and orientation. The behavior of male bots simulated that of a male straight user consuming *Bisexual Male* content. Similarly, female bot behavior simulated straight

³⁵The elbow method is an established graphic method for determining the optimal number of clusters in clustering analysis (Tibshirani et al., 2001). It is based on plotting the average within-cluster distance for multiple numbers of clusters and identifying the line’s inflection point, i.e. the “elbow” created by the line’s change in slope. For a typical application of the method in the social sciences refer to Garip, 2012).

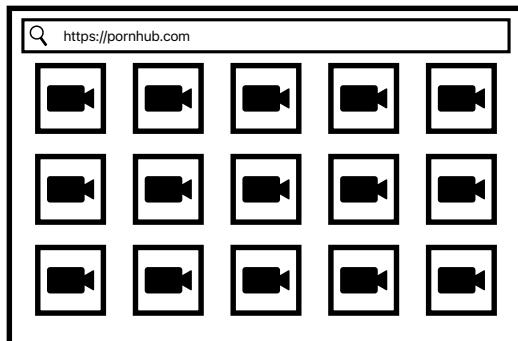


Figure 1: Schematic depiction of a video recommendation web-page. Video thumbnails appear as elements in a grid and users can click on a thumbnail to watch the respective video. The same video recommendation page is loaded across different parts of the platform, e.g., the “Hot videos in your Country” section (homepage), the “Recommended videos” section, or after selecting a certain category using a drop-down menu. The user interface largely resembles conventional video-sharing platforms such as YouTube.

female users watching *Lesbian* pornography.

Note that these categories are not analogous. The goal was to observe effects based on more nuanced activity, and thus, the *Bisexual Male* category was chosen for men. In any case, *Gay* videos do not frequently appear on Pornhub, as also shown by the analysis presented below³⁶. However, there is no corresponding category for *Bisexual Male* in women. As a result, the *Lesbian* category was used as the most closely related category to *Bisexual Male* for women users.

Finally, after I simulate each bot’s viewing of a seemingly non-normative video for their respective gender/sexual orientation, their recommendations are re-collected. The new recommendations are compared with the prior recommendations both within and across groups. The clusters discovered by the the word embedding model are used in the comparison as a content map of video category meaning.

³⁶More accurately, “Gay” videos do not appear on the “regular” or “conventional” Pornhub platform targeted to a straight audience (<https://www.pornhub.com>). This should be distinguished from the similar *Gay* version of the platform (<https://www.pornhub.com/gay>).

4 Results

4.1 Content Analysis

4.1.1 Term frequency analysis

Since I am interested in analyzing and understanding the platform’s content, it is worth exploring the most prevalent as well as the least popular content (Figure 2). The “rare” content appears to be directly related to male gay porn, with categories like *Gay* and *Twink*³⁷ appearing rarely on our random sample of videos. Also note that despite the alleged increasing popularity of *Hentai* (Pornhub, 2022a), few randomly selected videos appear under this category. On the other hand, popular categories include “placeholder” categories like *HD Porn*. Additionally, many videos seem connected with what could be considered mainstream pornography (e.g., *POV*³⁸). Some categories are targeted to a predominantly male audience reflecting the platform’s user base. For example, in 2023, 71% of the platform’s visitors in the United States were men (Pornhub, 2023b). The proportion of female users is also generally low worldwide, with 36% female visitors in 2023 (Pornhub, 2023b). A full list of category counts and their frequencies can be found in Table A3 of the Appendix.

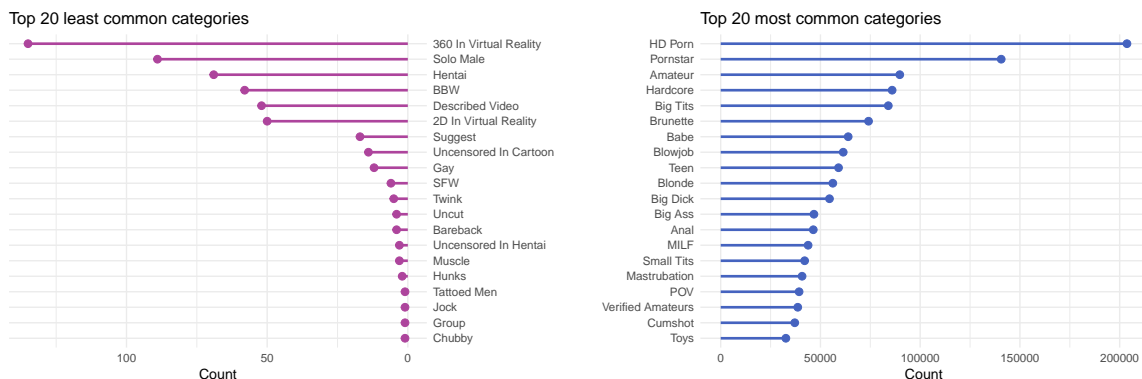


Figure 2: Categories with the fewer (left - pink) and most (right - blue) counts across the random sample of videos.

³⁷ *Twink* is a gay slang term usually referring to young gay or bisexual men with a thin body and little or no body hair, while recent conceptualizations of the term also include a passive anal sex role, colloquially known as “bottom” (Vytņiorgu, 2024).

³⁸ Acronym for Point of View. A particular style of mainstream pornography with videos shot with a wide camera angle from the point of view of the male performer.

4.1.2 Category trends

I also explore how these category counts change over time. For that, I selected a set of categories that will also appear in the rest of the analysis. Trends can be seen in Figure 3. Note that mainstream categories such as *Teen*, which also connect to “preying on young women” (Thompson, 2019) ³⁹, show a continuously increasing trend. The *Masturbation* category also shows an increasing trend at a lower rate. *Lesbian* videos, which are closely related to the male gaze, show somewhat more stable trends. *Transgender* video trends appear more unstable with multiple peaks and valleys, while it is important to mention that transgender videos appear exclusively after 2013. *Bisexual Male* videos appear earlier (2012) and constantly consist of the minority of videos posted on the site.

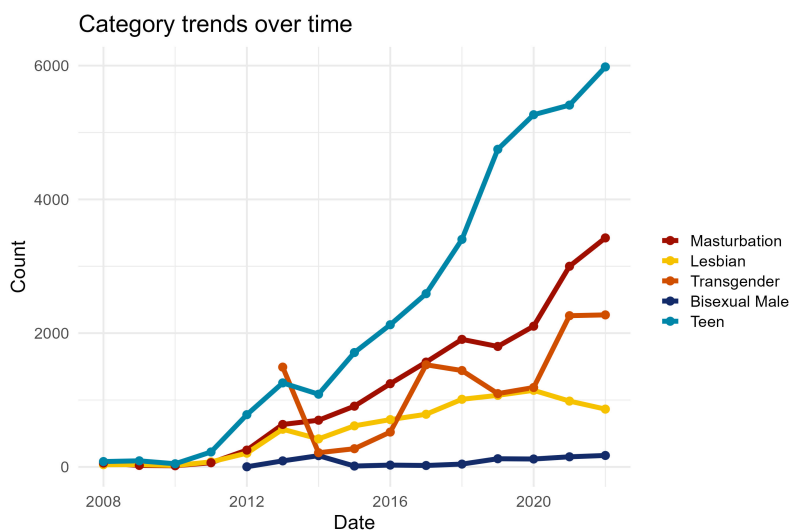


Figure 3: Category trends over time. Each point corresponds to the total number of videos belonging to the respective category for that year.

³⁹The concept of “preying on young women” is introduced by Stoya, pornographic actress, in an interview by Andrew Thomson, a Components researcher, (2019). The performer refers to the popularity of age-difference-related porn in which younger women are having intercourse with older men, often in accordance with an additional power dynamic such as a step-family relation (Thompson, 2019).

4.1.3 Semantic Network

Although aggregated category counts and temporal trends give us a sense of platform content, it is essential to understand category semantics analytically, i.e. what do all these categories really mean? As a first step towards understanding the platform meaning space, I create a semantic network of category co-occurrences and subsequently apply the Louvain community detection algorithm to group categories based on their connections within the graph.

The network captures some fundamental distinctions; for example, all the Transgender-related categories are part of the same community. We can also see that particular content niches are disconnected from the rest of the graph; examples include *Hentai*, *Cartoon*, or *Gay*. In particular, note that the Gay-related categories appear to be part of a relatively disconnected community from the rest of the graph. Overall, however, the graph appears to be heavily connected.

Following Mazières et al., 2014, I also explored the relation between centrality measures (measures of category importance) and category counts seen in Figure 5. Due to the highly connected structure of the network, little to no information can be extracted from these visualizations. The only finding is that categories such as *Hentai*, *Bareback* and *Gay* appear in the “unpopular” or “rare” categories, which was also evident from the network visualization, as well as expected by the norms of Pornhub’s straight version of the website.

An important distinction must be made between *Lesbian* and *Gay* categories, with the latter belonging to what Mazières et al., 2014 describe as “niche” categories; *Gay* videos are separated from the rest of the content by platform design. This is the version of the website targeted at a straight audience, yet *Lesbian* is part of the mainstream categories, highlighting the significance of the male gaze in pornographic representations.

Despite some informative results, the semantic network largely fails to capture nuanced meaning between categories due to its high connectivity. The vast majority of pornographic content is grouped around mainstream content.

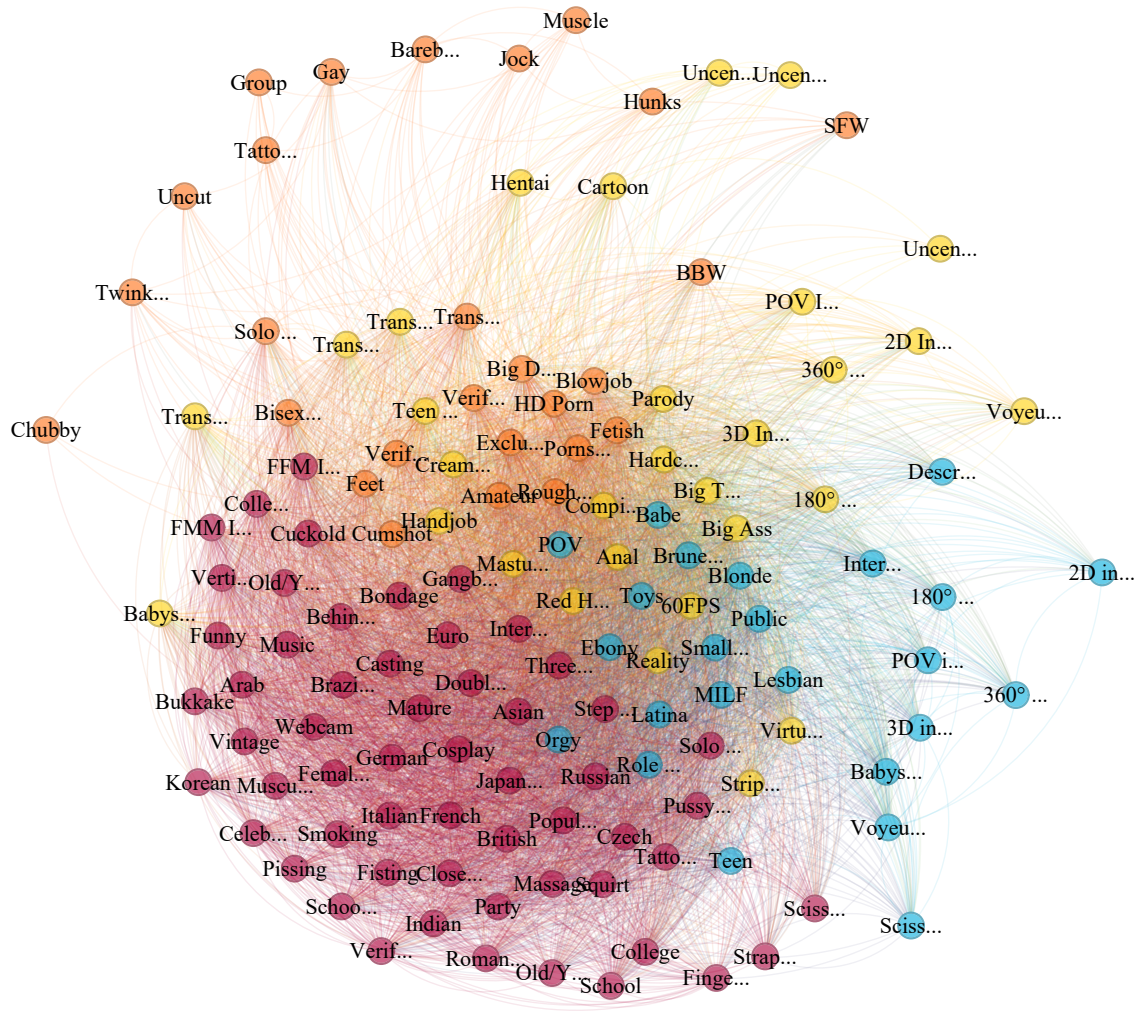


Figure 4: Semantic network of category co-occurrences. Categories are clustered on four different communities based on the Louvain algorithm. The resulting graph is densely connected, reducing its significance and usefulness as an analytical tool.

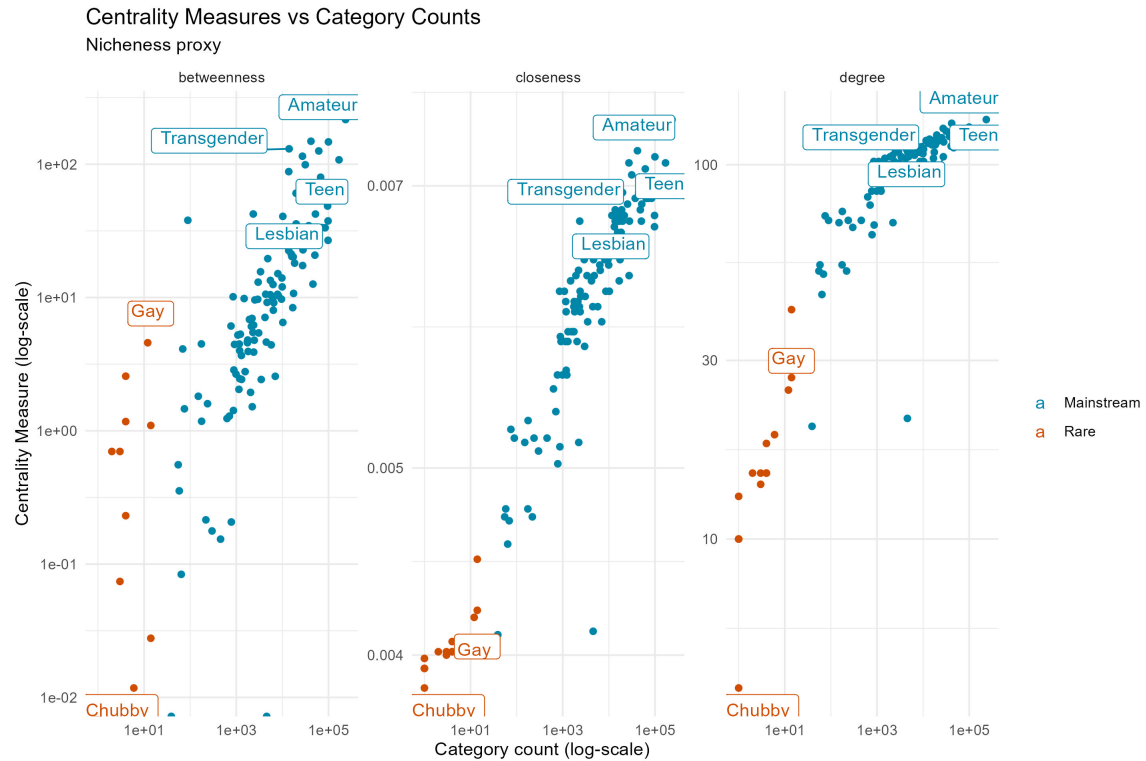


Figure 5: Popularity measures against category counts used as a “niceness” proxy, following Mazières et al., 2014. This plot is based on the semantic network of Figure 4. From left to right: betweenness, closeness, and node degree centrality measures. Apart from distinguishing rare (red) from popular (blue) categories (in agreement with the simple term frequency analysis of Figure 2), the insights we extract from this analysis are particularly limited.

4.2 Word embeddings

Since conventional content analysis methods, such as word frequency counts and semantic networks, fail to capture some nuances in category meaning, I employ word embeddings of video categories. I subsequently apply Principal Component Analysis (PCA) to the embedding vectors to reduce the meaning space into a 7-D principal component space. The number of components is determined using the scree plot method (Figure A3). The first seven principal components are then used as the input in a K-means clustering algorithm. The number of clusters is determined using the mean within-cluster distance as an evaluation metric and the Elbow method (Figure A4).

Based on these groupings, I then interpret the meaning space. Note the two levels of semiotic interpretation present in this analysis. First, category semantics are extracted directly through the word embedding model, treating video categories as isolated from English. Subsequently, the reduced meaning space is interpreted with the aid of English semantics. This approach to interpreting video categories is preferred to directly applying English semantics, as it guards the analysis from the researcher’s biases relatively well and should mitigate positionality concerns.

As a result, videos seem to be clustered in the following meaning clusters (Figure 6). :

- *Masturbation*: a cluster of videos belonging to categories like *Solo Female*, *Toys*, *Webcam*, and *Masturbation*, in which female actors are depicted masturbating.
- *Transgender*: a cluster of videos belonging to categories like *Trans Male In Transgender*, *Trans with Girl In Transgender*, *Trans with Guy In Transgender*, and *Transgender*.
- *Lesbian*: a cluster of videos belonging to categories like *Fingering*, *Scissoring in Lesbian*.⁴⁰, *Strap-on* ⁴¹, and *Lesbian*

⁴⁰*Scissoring* refers to a particular sex position in which partners sit or lie down crossing their legs and rubbing their genitals. It is a position generally attributed to lesbian sex, with mainstream films depicting scissoring acts such as Abdellatif Kechiche’s *Blue Is the Warmest Color*.

⁴¹*Strap-on* refers to a sexual toy, usually a plastic dildo imitating a penis, containing straps that can be

- *Mainstream/Men/Racial*: a cluster of videos belonging to both mainstream categories, categories depicting men and their body parts, as well as racial identities.
- *Mainstream/Women/White*: a cluster of videos belonging to both mainstream categories, categories referring to or targeted at women, as well as categories potentially signifying a white racial identity.
- *Orgy/Bisexual-Male*: a cluster of videos belonging to categories like *FMM in Threesome*⁴², *Muscular Men*, *Anal*.
- *Virtual Reality*: categories connected to VR pornography.

The entire set of categories belonging to each cluster can be seen in Tables A4 (*Masturbation*), A5 (*Transgender*), A6 (*Lesbian*), A7 (*Mainstream/Men/Racial*), A8 (*Mainstream/Women/White*), A9 (*Orgy/Bisexual-Male*), and A10 (*Virtual Reality*).

The two Mainstream clusters seem to distinguish themselves along multiple axes of meaning. Firstly, the *Mainstream/Women/White* cluster associates exclusively categories around female body parts like (*Big Tits*, *Small Tits*), hair color (*Blonde*, *Brunette*, *Red Head*), categories depicting women (*MILF*⁴³, *Tattooed Women*, *Babysitter*) or involving women (*Step Fantasy*⁴⁴, *Scissoring*, *FFM In Threesome*), fetishes (*BBW*⁴⁵, *Old/Young*, *Bondage*⁴⁶), categories that could be associated with soft porn (*Massage*, *Romantic*), and what could partly be attributed to white racial signifiers (*Blonde*, *Red Head*, *Euro*, *Czech*), as well as the *Popular With Women* category. Therefore, this cluster involves mainstream categories depicting mostly white women, referring to women, or categories tar-

attached on the wearer’s lap, i.e. it is ”strapped” on, hence its name. The toy is used in homosexual sex to penetrate either the vagina and/or the anus. At the same time, it is also used in hetero sex in which usually a woman penetrates their partner’s anus in an act called ”Pegging”.

⁴²This refers to Female-Male-Male Threesome, in which a woman and two men participate in an orgy. An alternative version is the *FFM Threesome* that contains two women and a man

⁴³Acronym for *Mother I’d Like to Fuck*, usually referring to a sexually attractive middle-aged woman

⁴⁴The *Step Fantasy* category refers to pornography depicting sexual acts of family incest between step-mothers, step-fathers, step-daughters, step-sons, step-sisters, and step-brothers in their respective heterosexual combinations (e.g. a step-mother and a step-son).

⁴⁵BBW stands for Big Beautiful Woman and is usually referring to overweight female bodies

⁴⁶*Bondage* refers to sexual acts involving tying and/or restraining the movement or one’s partner with ropes, leather straps, toys (e.g. gag balls), etc.

geted to women. That is the reason why I decided to name the cluster accordingly *Mainstream/Women/White*.

On the other hand, the *Mainstream/Men/Racial* cluster includes categories around different ethnicities and racial groups (*Asian, Latina, Indian, Arab, Ebony, French, Russian, German, Italian, Korean, Japanese*)⁴⁷, male body parts and ejaculation (*Blowjob, Handjob, Big Dick, Cumshot, Creampie*⁴⁸), the *dérrière* (*Big Ass, Anal*), depicting men (*Muscular Men, Solo Male*) alongside mainstream categories like (*Amateur, POV, Compilation*) and fetishes (*Hentai, Rough Sex, Public*). This cluster thus contains some mainstream categories, categories depicting men and/or their body parts, and racial/ethnic groups, and hence, it is named accordingly *Mainstream/Men/Racial*.

Note that similarity measures between the 100-D embedding space and the reduced 7-D PCA space match excellently, validating that the reduced space adequately captures the semantic relations of the entire embedding space. Categories cluster consistently at roughly the same five meaning axes. Clustering either at the 100-D embedding space or the 7-D reduced meaning space captures the most significant semantic relations among categories, as seen in the comparison of vector orthogonality presented in Table 2.

Category-1	Category-2	Cosine Similarity 100-D	Cosine Similarity 7-D
Trans With Girl In Transgender	Trans With Guy In Transgender	0.659370	0.920686
Trans Male In Transgender	Trans With Girl In Transgender	0.609182	0.723590
Trans Male In Transgender	Transgender	0.408467	0.627754
Trans With Girl In Transgender	Transgender	0.272695	0.538684
Trans Male In Transgender	Trans With Guy In Transgender	0.219868	0.520429
Transgender	Trans With Guy In Transgender	0.175653	0.503699

Table 2: Comparison of cosine similarities in the embedding (100-D) and reduced (7-D) space for transgender video categories. The similarity measure achieves a perfect match in both spaces; thus, the reduced space adequately captures semantic relations.

⁴⁷The categories belonging to this cluster include predominantly white ethnic groups such as *German* or *Russian*, alongside non-white ethnic groups such as *Indian*. In addition, some categories include a racial encoding excluding whiteness, e.g. *Ebony*.

⁴⁸*Creampie* refers to the act of ejaculating inside the anus or the vagina and letting the semen become visible dripping out after ejaculation.

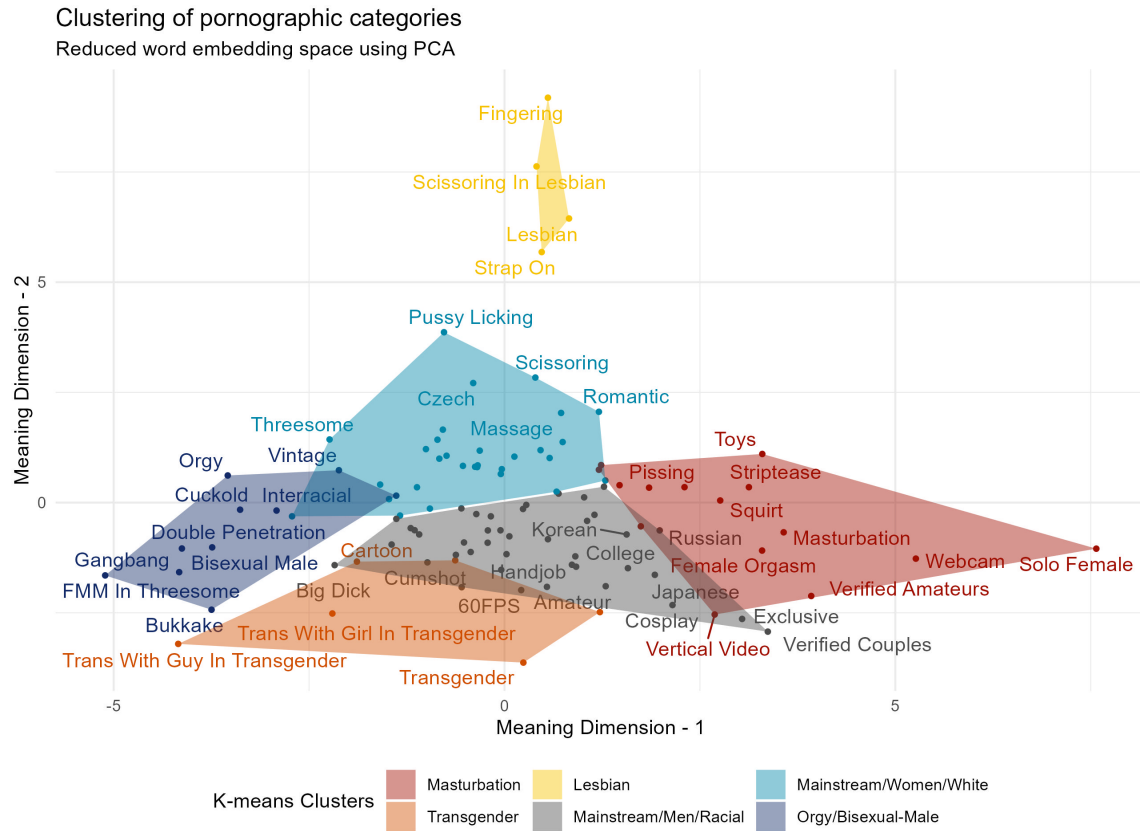


Figure 6: Reduced meaning space in 2-dimensions using PCA. Seven clusters of meaning are discovered using a k-means clustering method. The categories belonging to the Virtual Reality cluster are excluded for visual clarity.

4.3 Online experiments

The understanding of platform video category meaning can be expanded to other studies of algorithmic bias. As mentioned, I conducted a small-scale experiment using bots simulating female and male users of straight sexual orientation. The recommendation patterns before (prior treatment) and after (post-treatment) watching a non-normative video (*Lesbian* and *Bisexual Male* videos for female and male simulated users, respectively) are compared using the clusters of category meanings discovered by the word embeddings.

Here, the clusters of meaning are treated as a “map” of the meaning space, in the sense that clusters of meanings define areas on the 2-D plane defined by the first two principal components ($PCA-1$, $PCA-2$)⁴⁹. For that, the categories corresponding to recommended videos are juxtaposed on the meaning clusters as seen in Figure 7 (for comparisons of recommendation patterns across different PCA-dimensions, please see the Appendix).

Findings include little to no variation of recommendation patterns prior to treatment both within and across gender groups (Figure 7). Surprisingly, video recommendations before treatment for male bot users include at least one video belonging to the Transgender category (mainly a *Futanari* video⁵⁰). *Bisexual Male* videos appear on male recommendations exclusively post-treatment, while a significant shift of recommendation patterns towards *Lesbian* videos is observed for female bots.

Hence, Pornhub’s recommendation system seems to offer heteronormative initial recommendations for women but not men. While a shift of recommendation patterns towards non-normative videos is present in both gender groups, the shift is more significant for women than men.

⁴⁹While the 3-D reduced space representation captures more information regarding the meaning clusters, the 2-D version is chosen for simplicity.

⁵⁰*Futanari* refers to fictional characters that are otherwise female looking but have both female and male genitalia (Schuegraf and Tillmann, 2017). Futunari pornography is not, however, exclusively animated; edited videos of female pornstars with augmented digital penises (added digitally using special effects) also exist.

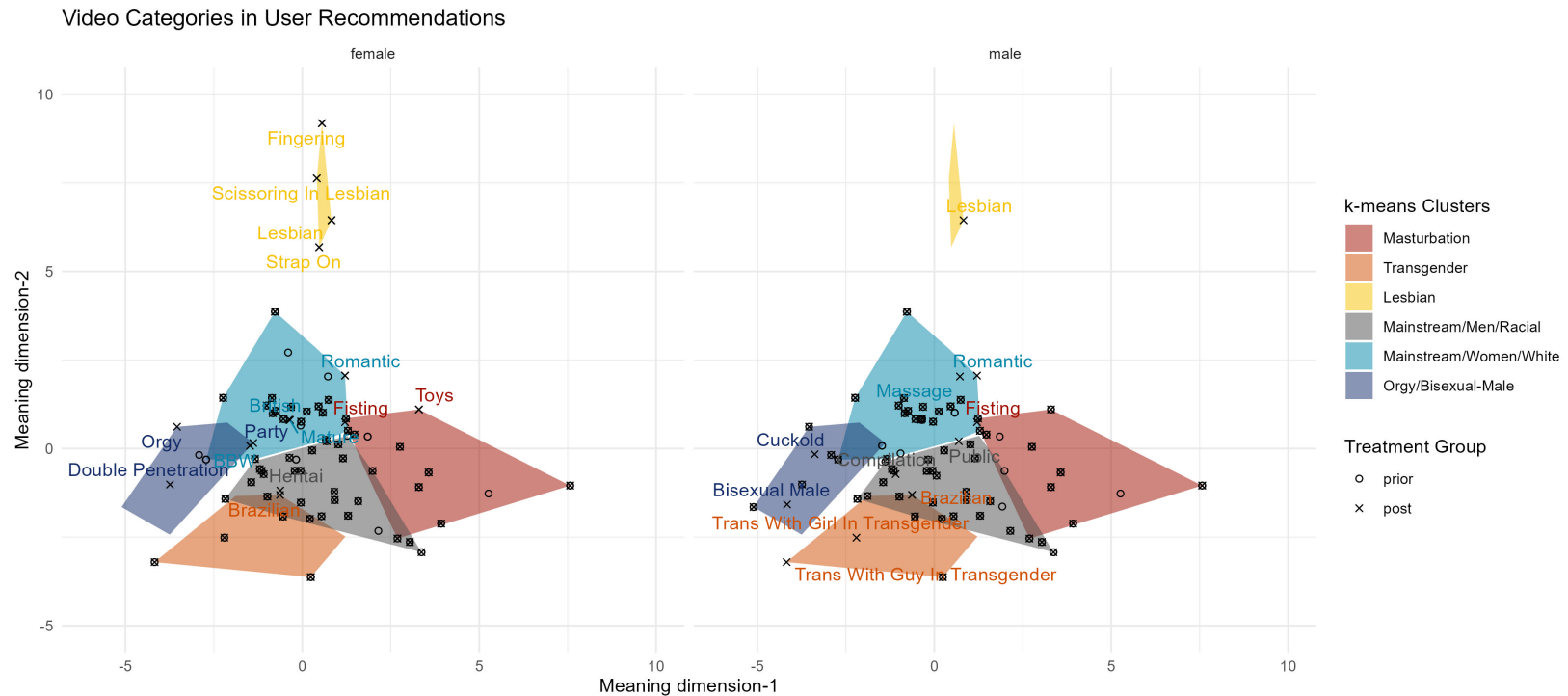


Figure 7: Recommendation patterns (categories of recommended videos) presented as black points (circles -o- for prior-to-treatment categories and crosses -x- for post-treatment categories), juxtaposed on the areas belonging to the six clusters of meaning for pornographic videos discovered in Figure 6. On the left, categories recommended for female bots are presented, whereas the graph on the right displays recommendations for male bots. Note that the experimental data are aggregated over the whole sample of users, as little to no variation of recommendation patterns is found within users of the same gender. The visible category names on the graph correspond to categories present solely in the video recommendations post-treatment.

5 Discussion

5.1 Methodological notes

The introduction of word embedding techniques into the analysis of pornographic videos proves particularly insightful and provides a novel, critical methodology for understanding platform categorization. Pornhub stands as one possible case study for the study of platform category meaning in online spaces.

The use of word embeddings to approximate category semantics is novel. The basic high-order assumption of the Word2Vec model (i.e. semantically similar words appear in similar contexts) is retained here despite the absence of a syntactical or grammatical structure (i.e. categories are simply a list of tokens). To grasp the totality of a video’s categorical context, Word2Vec’s window parameter is calibrated based on the distribution of the number of categories per video (Figure A1). This seems to be a simple yet effective way to deal with the absence of structure, as the model understands context in broad terms: the whole list of words is essentially each category’s context.

However, an absence of syntactical structure does not mean a lack of hierarchy. For example, the *Transgender* category acts as a placeholder for the other Transgender-related categories (*Trans with Guy in Transgender*⁵¹, *Trans with Girl in Transgender*⁵², *Trans Male in Transgender*⁵³). This hierarchy exists on the platform’s user interface design when users navigate the category sidebar. Nonetheless, the embedding model presented here deliberately does not consider hierarchical structures. Certain categories might share a hierarchical structure within the wider platform infrastructure. However, they co-occur on the same level under a video’s description (as part of the video category list), concealing the underlying hierarchy when users view a video. Therefore, for our purposes, categorical “ontology” is flat. The fact that, for example, Transgender-related categories cluster around their placeholder category by the model further validates the embedding approach. Results

⁵¹Pornography depicting a Trans woman having sex with a Cis man.

⁵²Pornography depicting a Trans woman having sex with a Cis woman.

⁵³Pornography depicting Trans men.

indicate that the model can extract meaning from category co-occurrences without grammatical, syntactical, or hierarchical structures. Further research needs to be conducted to benchmark this behavior.

The approach offered here presents a tool for effectively engaging with categorization in platforms around pornography and, by extension, cultural production. The relevant emerging literature of platform studies mainly rests on directly interpreting online spaces (Nieborg and Poell, 2018), independently interpreting categorization through term frequency tables (Stegeman et al., 2023) or conventional quantitative semantic analysis (Mazières et al., 2014). Despite the fruitful previous applications of such techniques, my studies of conventional content analysis fail to capture the nuance of meaning in pornographic categories unless this meaning is imported by the researcher conducting the study⁵⁴. Treating categories as linguistic structures mapped to an embedding space and isolated from English semantics or direct semiotic interpretations enables the extraction of meaning based on in-category semantics rather than importing semiotic meaning from language or social perceptions. This analytical tool helps insulate relevant research from positionality concerns and provides a detailed and nuanced understanding of the social norms and hierarchies arising in the platform’s content.

5.2 Category meaning space

As discussed above, the category space model used 100 dimensions in order to allow for sufficient spatial dimensions to capture nuances in category meaning. Results were compared across multiple dimensions and tuned based on the interpretability of the final representation. Out of the seven broad categories found (*Masturbation*, *Transgender*, *Lesbian*, *Mainstream/Men/Racial*, *Mainstream/Women/White*, *Orgy/Bisexual-Male*, and *Virtual Reality -VR-*), the VR cluster was excluded from the study as the least interesting. The rest of

⁵⁴On that note, we could, in theory, create an index or labels based on English semantics and social perceptions by simply looking at platform categories. For example, looking at the Solo Male and Masturbation categories, one could arrange them into a group of videos depicting masturbation. However, my abstract quantitative approach, using word embeddings, seems to propose that the Masturbation section refers mainly to female self-love rather than men jerking off on camera. Further walk-through exploration of these categories on Pornhub validates these insights.

the clusters capture semantic meaning in a particularly insightful and easily interpretable way. Word similarities were found to be consistent when the dimensions of the space were increased up to three hundred. Further, decreasing the dimensions of the embedding space below a hundred did not enable the model to capture meaning complexity. In general, the dimensions need to be adjusted according to the use-case scenario since a consistent way to treat dimensions of the space as a hyperparameter is still ongoing in embedding research (for example, see the work of Patel and Bhattacharyya, 2017 and Yin and Shen, 2018).

The approach presented also provides an analytically consistent way of treating dimensionality reduction, a common technique in extracting interpretable results from word embeddings (Mu et al., 2018; Raunak et al., 2019), and clustering using a technique of graphical optimization (Elbow method). Dimensionality reduction and feature extraction provide sufficiently interpretable results, while the methodology offers a consistent way of applying word embedding models to platform categorization research. The process can be further validated by comparing term similarities across the complete 100-D and 7-D reduced space for categories critical to the analysis (refer, for example, to the *Transgender* categories comparison presented in Table 2).

Lesbian The *Lesbian* videos cluster consists of the most strongly connected categories. The categories *Fingering*, *Lesbian*, *Scissoring In Lesbian*, and *Strap On* consistently cluster together across multiple reduced space dimensions. This means the categories are tightly knit and have a clear meaning distinct from the rest of the category space. Lesbian videos appear spatially disjoint and distant from the rest of the categories in the reduced meaning space; Lesbian videos are essentially a world of their own.

Transgender *Transgender* videos seem to group consistently in the same cluster as well. One might initially question the existence of the *Cartoon* category in the Transgender space. However, a walk-through examination of Pornhub will uncover that Cartoon videos often refer to cartoonish or animated depictions of transgender people (mainly transgender women) having sex. The extent to which animated transgender pornography demeans this

social group, oversexualizing the trans body or treating it as a fetish fantasy warrants further exploration.

The existence of *Brazilian* among the *Transgender* cluster is no surprise either. According to Pornhub’s yearly reports, Brazilians’ interest in Transgender pornography has been consistently high in recent years, the category being the most viewed on-site for 2022 and the third most popular in 2023, increasing in popularity by 68% (Pornhub, 2022a, 2023b).

Orgy/Bisexual-Male The *Orgy/Bisexual-Male* cluster is the first in which we can see a blur between mainstream heterosexual pornography and seemingly non-normative sex acts. The cluster functions both as a space for group sex categories and also as an expression of bisexuality. This is self-explanatory as for a male performer to be seen as bisexual, they need to participate in group sex with both a man and a woman (hence the existence of *FMM In Threesome* on the same cluster, too⁵⁵).

Another finding is that the *Interracial* category is part of this particular cluster. This adds to our knowledge of race fetishization in pornographic spaces (Neal, 2013). The fact that *Interracial* videos appear primarily in the context of group sex amplifies concerns around racism depictions in pornography (Felkins, 2015), as group sex in mainstream pornography is often seen as dehumanizing and degrading. The presence of interracial sex in the context of orgies potentially symbolizes the perceived dominance of the white over the black body in mainstream pornographic representation.

Masturbation *Masturbation* videos refer exclusively to female self-love. The embedding model captures excellently nuances in meaning, with categories like *Toys*, *Pissing*, *Squirt*, *Webcam*, and *Solo Female* all clustering around the same area in the meaning space. Interestingly enough, the concepts of feet fetish in particular and fetish in general are associated with these categories. The presence of *Feet* might be attributed to multiple female masturbation videos showing the performer’s feet because of the camera angle, as discovered

⁵⁵An *FMM In Threesome* is a sufficient but not necessary condition for content bisexuality; there are both straight and bisexual *FMM In Threesome* videos.

by a simple walk-through on Pornhub. Nonetheless, the association of smoking with masturbation videos (through the *Smoking* category) needs to be further explored, as no clear association between the two can be extracted by exploring Pornhub qualitatively; the extent to which this connection is an artifact of dimensionality reduction also needs to be further studied.

Mainstream clusters The two mainstream category clusters are the hardest to interpret. The two main meaning axes distinguishing between the two are the representation of male and female bodies, while blurred racial distinctions also exist. A “White” category might not exist in Pornhub, in agreement with what Saunders, 2020 suggests and I hinted at earlier. However, we see that the concept of whiteness might be encoded in specific categories signifying it (*Blonde, Red Head, Euro, etc.*); in that sense, a “White” category, or more accurately, a group of “White” categories, does exist. Thus, a nuanced meaning behind categories is uncovered. Categories pointing to hair color do not simply refer to hair but carry the concept of whiteness, much in agreement with Stegeman et al., 2023 findings regarding webcam sex platforms. In addition, the presence of female-related categories within a cluster of categories referring to white bodies is not a coincidence, as white bodies are also the most desired bodies within the online sex-work space (Jones, 2015, 2020; Stegeman et al., 2023).

The co-existence of male-body parts with racially-coded categories is also not coincidental. These findings reinforce concerns around the fetishization of black penises in white fantasies (Marriott, 1996) or the orientalization/fetishization of Asian female bodies (Pires, 2024). The surprising co-existence of *Anal* with racially-coded categories must also be put into a framework of connecting blackness with “abnormal” or even “exotic” pleasures (Nash, 2014). Overall, content across the two clusters could be read as symbolizing the domination of male/white bodies over female bodies, which are at times desired (White bodies) or exoticized (Black and Asian bodies). The mere clustering of all non-white ethnic and racial groups under the same cluster points to a fetishization element over the non-white body.

Category Imbalance The imbalance of content representing homosexual intercourse is also striking. The imbalance is a baseline for the platform, as videos corresponding to the *Gay* category are largely missing from the straight version of Pornhub, exhibited by the low counts of gay videos in our random sample. This positions Pornhub as a space curated almost exclusively for a straight men audience. At the same time, it spurs concerns about representation, especially since many straight women prefer watching male gay pornography (Bell et al., 2019; Neville, 2015). In addition, the widespread availability of *Lesbian* videos on the platform corroborates conversations around the portrayal of lesbian intercourse through the male gaze (Neville, 2015; Wolfe and Roripaugh, 2006).

5.3 Experimentation

Video recommendations do not seem to vary across the male/female gender types when the account has been created, and the platform has not yet collected any data on-site other than the self-declared gender and sexual identity of the user’s profile. This, on the one hand, makes us classify the recommendation system as biased towards heteronormative content or, at the very least, not accounting for gender-based differences in the sexual preferences of straight men and women. Nonetheless, a closer look into the categories recommended to straight users complicates the situation: Both male and female users are consistently fed videos belonging to the *Transgender* category. The frequency of transgender-themed videos recommended to female users exceeds that of male users, who at most receive one transgender video suggestion (usually a *Futanari* video). In that sense, the initial recommendation patterns for both men and women do not solely feature heteronormative content, although neither receives recommendations for homosexual videos for their corresponding gender (*Gay* or *Bisexual* videos for men and *Lesbian* videos for women).

After the synthetic users watch a video contradicting their initially declared sexual orientation (*Lesbian* for women and *Bisexual-Male* for men), each user’s recommendation patterns change accordingly to encompass the new categories. Female synthetic users can easily bypass hetero-normative constraints on their original recommendations, such as the

absence of *Lesbian* videos. The recommendation system immediately picks up interest in *Lesbian* porn, suggesting multiple videos of the *Lesbian* cluster after watching just one relevant video.

The effect, however, is much more nuanced on men. The presence of transgender-themed videos might be strengthened, and simultaneously, more *Bisexual-Male* videos pop up, but the shift in recommendation patterns is less extreme. A plausible explanation is that the platform considers watching a bisexual male video by a straight man not as a strong indicator of preference for homosexual pornography. In that sense, sexual plasticity is assumed for men. In addition, the presence of transgender videos in prior-to-treatment male recommendations points to the existence of degrees of freedom in sexual desires for men. This might also be due to the perceived sexual identity of men sexually interested in transgender women. As Weinberg and Williams, 2010 point out, many men attracted by transgender women are identifying as straight⁵⁶. On the contrary, a straight woman watching a lesbian video signals to the platform a more substantial interest in homo-erotic love; female sexuality is perceived as more rigid by the system. Therefore, the platform's recommendation system, although not necessarily as biased as previously thought, does assume more sexual plasticity for men rather than women.

Despite the gender-based differences in recommendations and the system's behavior, the algorithm seems to adapt well to user preferences. In addition, assuming that recommendations are based on past user profiling, straight men engage in watching non-heteronormative content without significantly affecting their declared sexual identities. Overall, we could say that the algorithmic recommendation system enforces heteronormativity more on women than men and assumes a more rigid conceptualization of sexuality for women than men, while both groups achieve a shift of initial biases in recommendations.

Note that the extent to which the content recommended to straight women who want to watch lesbian pornography is acting towards bypassing heteronormativity is itself open to discussion, considering that a large part of mainstream Lesbian pornography is targeted at

⁵⁶Also note that transgender pornography recently assumed a position in mainstream porn culture, as pointed out by Pornhub's yearly reports (Pornhub, 2023b).

a male straight audience (Meehan, 2021). However, again, assuming that the recommendation system suggests videos based on profiling of past viewing habits, the platform seems to consider Lesbian pornography as not part of what a straight woman would watch. On the one hand, this exclusion of lesbian content could be interpreted as enforcing heteronormativity and delimiting straight female desire. On the other hand, it could be attributed to mainstream Lesbian pornographic content, which, as targeted to a male straight audience, deters female straight users from watching it.

Categorical imbalance on treatment Categories chosen as the “treatment” (*Lesbian* for female users and *Bisexual Male* for male users) are not balanced. I decided to simulate the treatment by making male straight bots watch a *Bisexual Male* video to capture a nuanced difference in the sexuality of a straight man attempting to watch content contradicting their declared identity on site. However, there is no analogous category for women (e.g., Bisexual-Female). Thus, the *Lesbian* category was chosen as the most closely analogous category to *Bisexual Male*. Note that the *Gay* category is eventually excluded from the analysis as it is not well represented on the site and, therefore, cannot be used for bots simulating male users. This observed categorical imbalance for homosexual content could potentially explain the absence of lesbian videos in female recommendations prior-to-treatment; lesbian pornography is largely intended to be consumed by straight men and possibly does not appeal to a female audience.

6 Conclusions and Future Trajectories

Word embeddings provide a powerful heuristic and analytical tool for uncovering the underlying semantics of categories in online platforms. Applying this technique to Pornhub, I demonstrate the presence of cultural patterns regarding representation, desire, and social hierarchies, as well as systemic ingrained biases in the platform’s algorithmic recommendation system. The extent to which heteronormativity unfolds within the site was also examined. Algorithmic systems and content seem to work closely together to partly reinforce and

partly challenge the prevailing norms on the site.

Future studies should incorporate large-scale studies of algorithmic bias based on the content mapping produced in this analysis. Increasing category statistics could also enable the model to better capture nuances in category semantics. In addition, the expansion of the work towards analyzing the *Gay* version of the platform, as well as other popular pornographic websites, is necessary to understand better how platforms categorize and curate their content. The end goal of this research trajectory would be to create an analytically rich, spatially and temporally fine-grained semantic representation of pornographic categories. Additionally, introducing word embedding analysis on tags could provide insights into semantic differences between top-down (categories) and bottom-up (tags) categorization processes.

Regarding the study of algorithmic bias in Pornhub, the need to examine different locations, genders, and sexual orientations is necessary to critically assess how content interacts with algorithmic systems to reinforce existing norms or challenge them.

I hope the method proposed here will be used in the study of pornographic content and broadly in the emergent field of critical Platform Studies to understand better and critique platforms that govern our lives, desires, identities, and needs. A semantic understanding of terms used across platforms is necessary in order to challenge this top-down imposed categorization; in accordance with Antisthenes' aphorism: "ἀρχὴ σοφίας ονομάτων ἐπίσχεψις"⁵⁷.

⁵⁷ "The investigation of the meaning of words is the beginning of wisdom".

A Appendix

180o In Virtual Reality	Bondage	Euro	Indian	Pissing	Solo Male	Verified Couples
2D In Virtual Reality	Brazilian	Exclusive	Interactive	Popular With Women	Squirt	Verified Models
360 In Virtual Reality	British	Feet	Interracial	Pornstar	Step Fantasy	Vertical Video
3D In Virtual Reality	Brunette	Female Orgasm	Italian	POV	Strap On	Vintage
60FPS	Bukkake	Fetish	Japanese	POV In Virtual Reality	Striptease	Virtual Reality
Amateur	Cartoon	FFM In Threesome	Jock	Public	Suggest	Voyeur In Virtual Reality
Anal	Casting	Fingering	Korean	Pussy licking	Tattooed men	Webcam
Arab	Celebrity	Fisting	Latina	Reality	Tattooed Women	
Asian	Chubby	FMM In Threesome	Lesbian	Red Head	Teen	
Babe	Closed Captions	French	Massage	Role Play	Threesome	
Babysitter	College	Funny	Masturbation	Romantic	Toys	
Bareback	Compilation	Gangbang	Mature	Rough Sex	Trans Male In Transgender	
BBW	Cosplay	Gay	MILF	Russian	Trans With Girl In Transgender	
Behind The Scenes	Creampie	German	Muscle	School	Trans With Guy In Transgender	
Big Ass	Cuckold	Group	Muscular Men	Scissoring	Transgender	
Big Dick	Cumshot	Handjob	Music	Scissoring In Lesbian	Twink	
Big Tits	Czech	Hardcore	Old/Young	SFW	Uncensored In Cartoon	
Bisexual Male	Described Video	HD Porn	Orgy	Small Tits	Uncensored In Hentai	
Blonde	Double Penetration	Hentai	Parody	Smoking	Uncut	
Blowjob	Ebony	Hunks	Party	Solo Female	Verified Amateurs	

Table A1: Video categories found in my randomly sampled collection of video metadata. The categories are sorted alphabetically across columns.

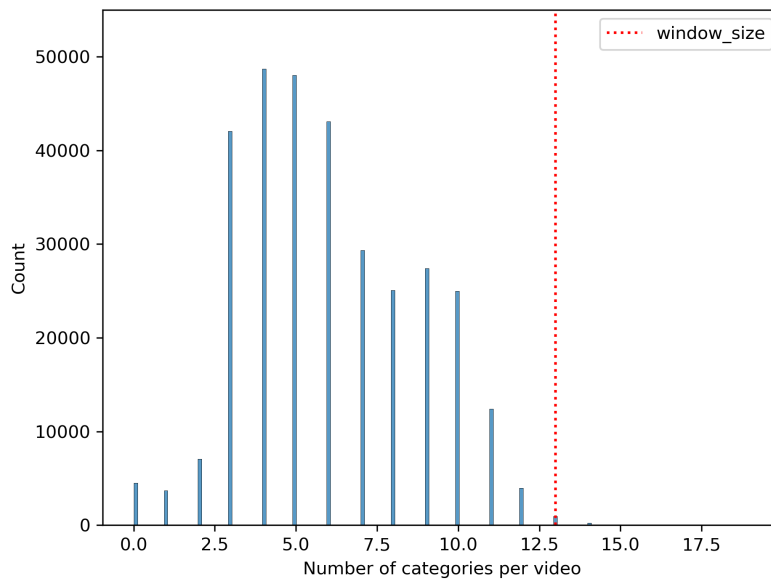


Figure A1: Distribution of the number of categories per video used to determine the window size parameter of the Word2vec model.

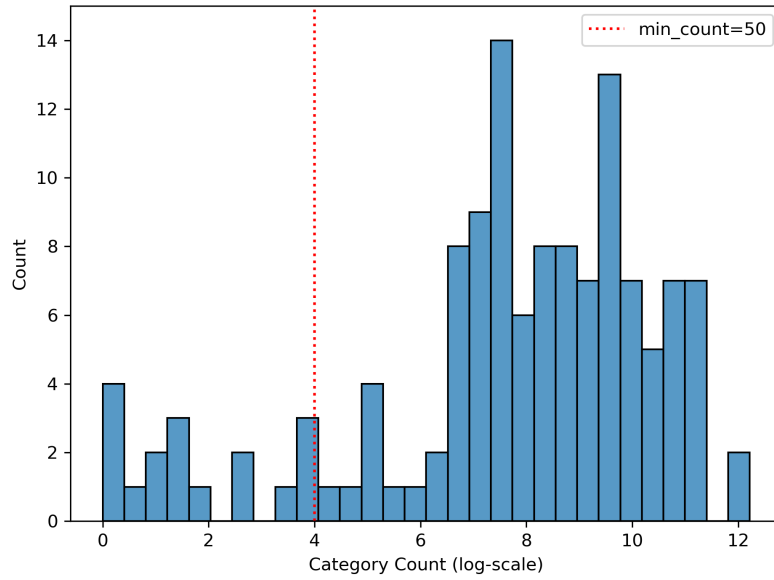


Figure A2: Distribution of category counts used to determine the min-count cutoff for the word embedding model.

Category	Count
Described Video	48
Uncensored In Cartoon	14
Gay	12
SFW	6
Twink	5
Bareback	4
Uncut	4
Muscle	3
Uncensored In Hentai	3
Hunks	2
Chubby	1
Group	1
Jock	1
Tattooed Men	1

Table A2: Rare/Unpopular categories of very few counts that are subsequently dropped from the word embedding model's input.

Category	Count	Frequency (%)
HD Porn	202707	63.0
Pornstar	144153	44.8
Amateur	89132	27.7
Teen	88165	27.4
Hardcore	85596	26.6
Big Tits	83633	26.0
Brunette	73831	23.0
Babe	63542	19.8
Blowjob	61190	19.0
Blonde	55961	17.4
Big Dick	54333	16.9
Big Ass	46561	14.5
Anal	46120	14.3
MILF	43563	13.5
Small Tits	41953	13.0
Masturbation	40530	12.6
POV	39090	12.2
Verified Amateurs	38537	12.0
Cumshot	37022	11.5
Toys	32492	10.1
Exclusive	27770	8.63
Fetish	24631	7.66
Threesome	24149	7.51
Lesbian	23298	7.24
Reality	22374	6.96
Interracial	18943	5.89
Rough Sex	18313	5.69

Creampie	17696	5.50
Asian	16618	5.17
Public	16335	5.08
Pussy Licking	15794	4.91
Solo Female	15561	4.84
Latina	15237	4.74
Ebony	14870	4.62
Step Fantasy	14487	4.50
Transgender	13989	4.35
Popular With Women	12730	3.96
Handjob	12064	3.75
Verified Models	12007	3.73
Red Head	11941	3.71
60FPS	11895	3.70
Compilation	10909	3.39
Female Orgasm	10167	3.16
Japanese	9134	2.84
Euro	8649	2.69
Mature	8604	2.68
Orgy	8583	2.67
Webcam	8023	2.49
Squirt	7333	2.28
Role Play	7057	2.19
Double Penetration	6880	2.14
Verified Couples	6758	2.10
Tattooed Women	6100	1.90
Russian	5722	1.78
Bondage	5719	1.78

Gangbang	5570	1.73
Old/Young	4939	1.54
Striptease	4785	1.49
Massage	4593	1.43
Romantic	4135	1.29
Casting	4052	1.26
German	3941	1.23
British	3939	1.22
Czech	3697	1.15
Vintage	3071	0.955
College	2999	0.932
Trans With Guy In Transgender	2978	0.926
Party	2686	0.835
Feet	2332	0.725
Cosplay	2324	0.723
Brazilian	2266	0.705
Fingering	2140	0.665
Pissing	2099	0.653
Fisting	2092	0.650
Virtual Reality	1983	0.617
Indian	1928	0.599
French	1855	0.577
Bukkake	1815	0.564
FFM In Threesome	1740	0.541
School	1718	0.534
Music	1693	0.526
Behind The Scenes	1668	0.519
Cuckold	1594	0.496

Muscular Men	1516	0.471
Italian	1366	0.425
FMM In Threesome	1303	0.405
Bisexual Male	1225	0.381
Vertical Video	1182	0.368
3D In Virtual Reality	1172	0.364
Celebrity	1106	0.344
180 In Virtual Reality	1075	0.334
Smoking	1042	0.324
Closed Captions	969	0.301
Arab	815	0.253
Funny	811	0.252
Babysitter	806	0.251
Scissoring In Lesbian	781	0.243
Strap On	769	0.239
Trans With Girl In Transgender	768	0.239
Parody	744	0.231
Korean	587	0.183
POV In Virtual Reality	561	0.174
Interactive	366	0.114
Trans Male In Transgender	298	0.0927
Voyeur In Virtual Reality	186	0.0578
Cartoon	175	0.0544
Scissoring	151	0.0469
360 In Virtual Reality	138	0.0429
Solo Male	89	0.0277
Hentai	69	0.0215
BBW	58	0.0180

2D In Virtual Reality	50	0.0155
Described Video	48	0.0149
Uncensored In Cartoon	14	0.00435
Gay	12	0.00373
SFW	6	0.00187
Twink	5	0.00155
Bareback	4	0.00124
Uncut	4	0.00124
Muscle	3	0.000933
Uncensored In Hentai	3	0.000933
Hunks	2	0.000622
Chubby	1	0.000311
Group	1	0.000311
Jock	1	0.000311
Tattooed Men	1	0.000311

Table A3: Complete category counts and frequencies of category occurrence.

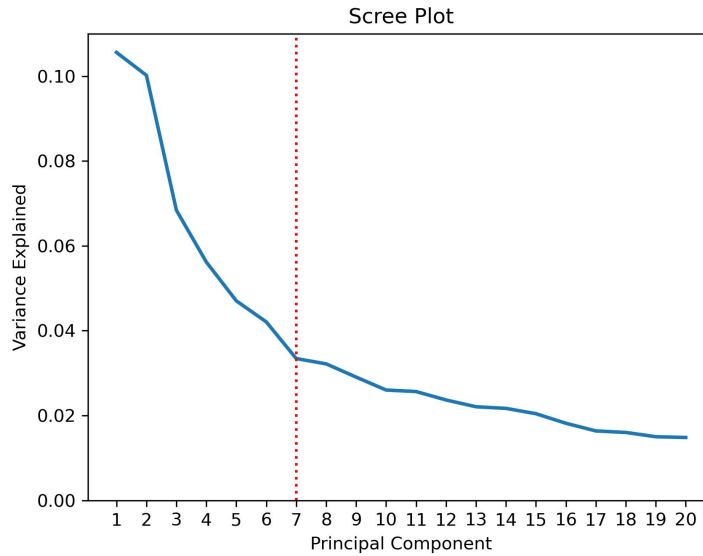


Figure A3: Scree plot to determine the number of principal components to keep while applying dimensionality reduction on the initial 100-D embedding space. The number of dimensions explaining the majority of data variance is determined at seven (red dotted line) by graphically identifying the line's inflection point using the Elbow Method.

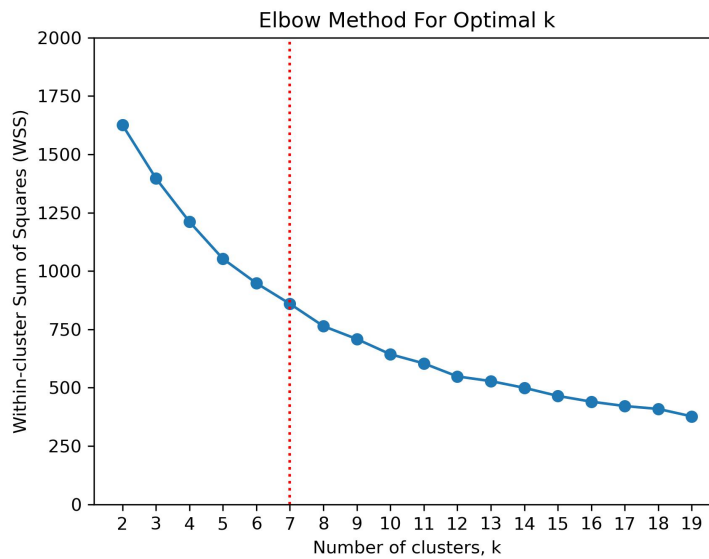


Figure A4: Determining the number of clusters using the Elbow Method, graphically identifying the line's inflection point (red dotted line). The number of meaningful clusters is thus identified as seven.

Feet	Female Orgasm	Fetish
Fisting	Masturbation	Pissing
Smoking	Solo Female	Squirt
Striptease	Toys	Verified Amateurs
Verified Models	Vertical Video	Webcam

Table A4: Categories belonging to the *Masturbation* cluster.

Brazilian	Cartoon
Trans Male In Transgender	Trans With Girl In Transgender
Trans With Guy In Transgender	Transgender

Table A5: Categories belonging to the *Transgender* cluster.

Fingering	Lesbian
Scissoring In Lesbian	Strap On

Table A6: Categories belonging to the *Lesbian* cluster.

60FPS	Amateur	Anal	Arab
Asian	Babe	Behind The Scenes	Big Ass
Big Dick	Blowjob	Celebrity	College
Compilation	Cosplay	Creampie	Cumshot
Ebony	Exclusive	French	Funny
German	HD Porn	Handjob	Hentai
Indian	Interactive	Italian	Japanese
Korean	Latina	Muscular Men	Music
POV	Parody	Public	Role Play
Rough Sex	Russian	School	Solo Male
Verified Couples			

Table A7: Categories belonging to the *Mainstream/Men/Racial* cluster.

BBW	Babysitter	Big Tits	Blonde
Bondage	British	Brunette	Casting
Closed Captions	Czech	Euro	FFM In Threesome
Hardcore	MILF	Massage	Mature
Old/Young	Popular With Women	Pornstar	Pussy Licking
Reality	Red Head	Romantic	Scissoring
Small Tits	Step Fantasy	Tattooed Women	Teen
Threesome			

Table A8: Categories belonging to the *Mainstream/Women/White* cluster.

Bisexual Male	Bukkake
Cuckold	Double Penetration
FMM In Threesome	Gangbang
Interracial	Orgy
Party	Vintage

Table A9: Categories belonging to the *Orgy/Bisexual-Male* cluster.

180o In Virtual Reality	2D In Virtual Reality
360o In Virtual Reality	3D In Virtual Reality
POV In Virtual Reality	Virtual Reality
Voyeur In Virtual Reality	

Table A10: Categories belonging to the *Virtual Reality* cluster.

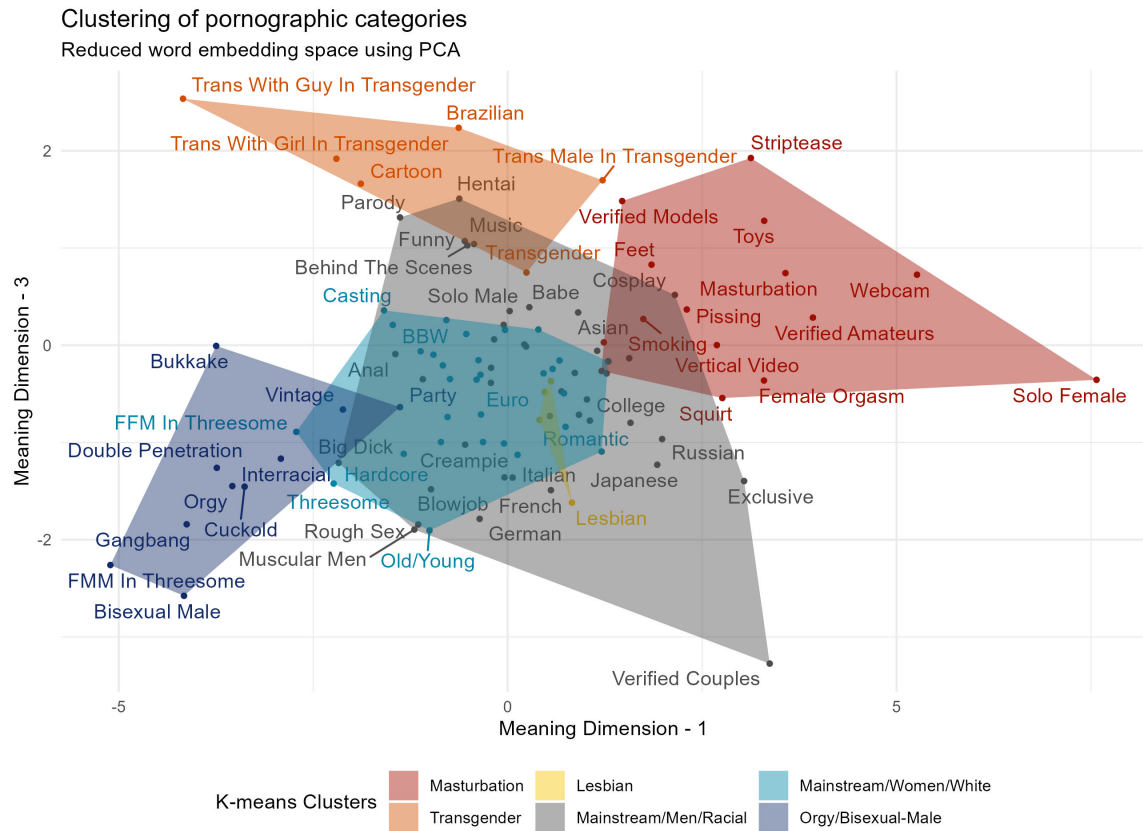


Figure A5: Reduced meaning space in 2-dimensions using PCA visualized on the plane of the first and third principal components ($PCA - 1 - PCA - 3$). Seven clusters of meaning are discovered using a k-means clustering method. The categories belonging to the Virtual Reality cluster are excluded for visual clarity.

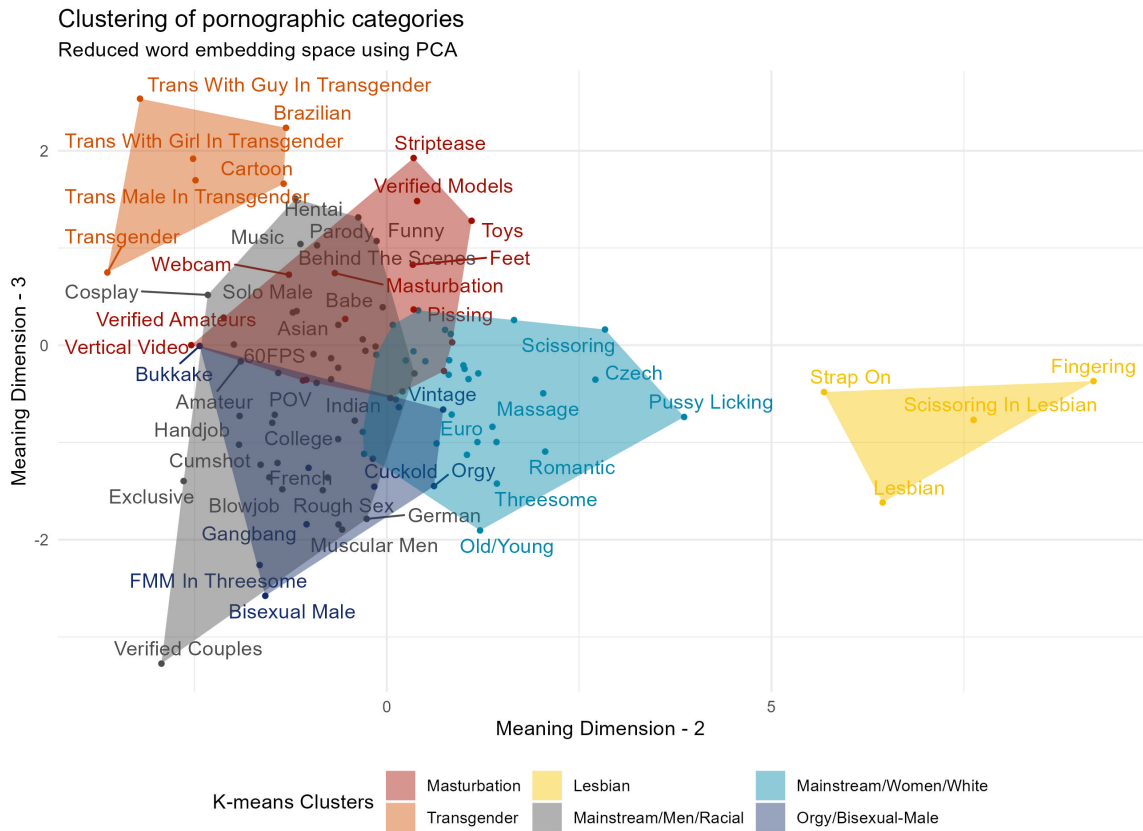


Figure A6: Reduced meaning space in 2-dimensions using PCA visualized on the plane of the second and third principal components ($PCA-2-PCA-3$). Seven clusters of meaning are discovered using a k-means clustering method. The categories belonging to the Virtual Reality cluster are excluded for visual clarity.

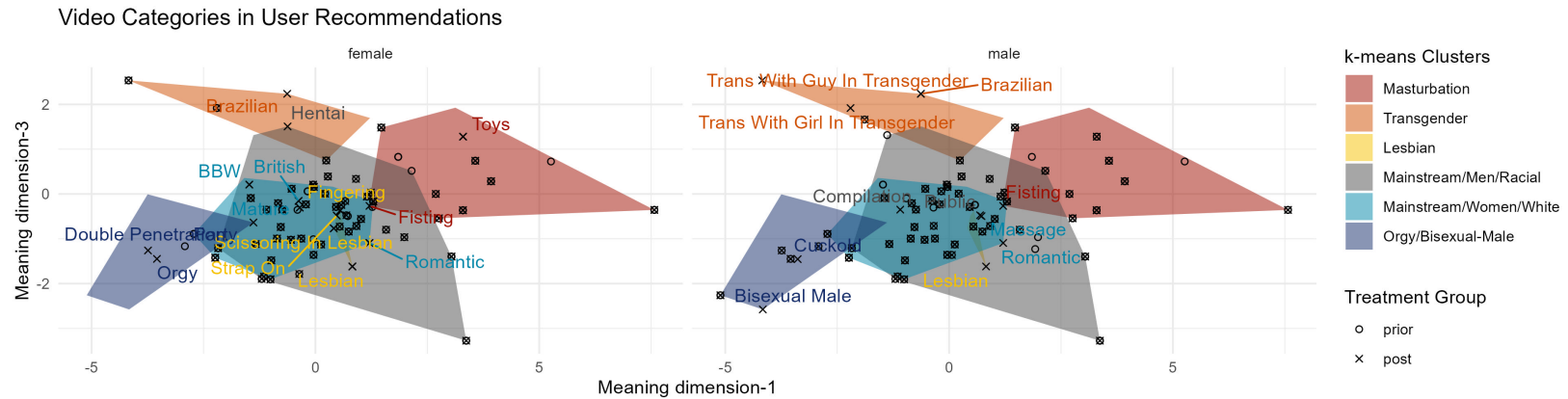


Figure A7: Recommendation patterns (categories of recommended videos) presented as black points (circles -o- for prior-to-treatment categories and crosses -x- for post-treatment categories), juxtaposed on the areas belonging to the six clusters of meaning for pornographic videos discovered in Figure A5. On the left, categories recommended for female bots are presented, whereas the graph on the right displays recommendations for male bots. Note that the experimental data are aggregated over the whole sample of users, as little to no variation of recommendation patterns is found within users of the same gender. The visible category names on the graph correspond to categories present solely in the video recommendations post-treatment.

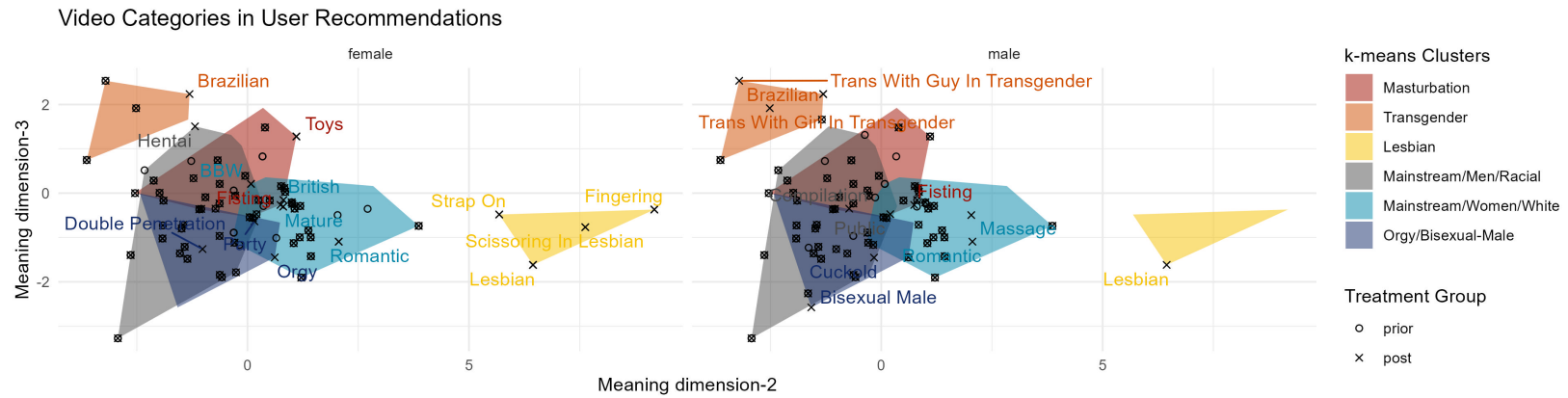


Figure A8: Recommendation patterns (categories of recommended videos) presented as black points (circles -o- for prior-to-treatment categories and crosses -x- for post-treatment categories), juxtaposed on the areas belonging to the six clusters of meaning for pornographic videos discovered in Figure A6. On the left, categories recommended for female bots are presented, whereas the graph on the right displays recommendations for male bots. Note that the experimental data are aggregated over the whole sample of users, as little to no variation of recommendation patterns is found within users of the same gender. The visible category names on the graph correspond to categories present solely in the video recommendations post-treatment.

References

- Agocha, V. B., Asencio, M., & Decena, C. U. (2014). Sexuality and culture. In *APA handbook of sexuality and psychology, Vol. 2: Contextual approaches* (pp. 183–228). American Psychological Association. <https://doi.org/10.1037/14194-006>
- Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The k-means Algorithm: A Comprehensive Survey and Performance Evaluation. *Electronics, 9*(8), 1295. <https://doi.org/10.3390/electronics9081295>
- Ainsworth-Masiello, R., & Evans, D. T. (2019). Expectations vs reality: In which ways might watching porn online, as male and female adolescents, contribute to poor emotional health? *Education and Health, 37*(4), 109–116. Retrieved March 31, 2024, from <http://sheu.org.uk/sheux/EH/eh374ram.pdf>
- Almeida, F., & Xexéo, G. (2023, May). Word Embeddings: A Survey. <https://doi.org/10.48550/arXiv.1901.09069>
- Aristotelis. (2011). *Kategoriai, Peri Ermineias* (P. Kalligas & A. Mpoukala, Eds.). Nisos.
- Attwood, F. (2002). Reading Porn: The Paradigm Shift in Pornography Research [Publisher: SAGE Publications Ltd]. *Sexualities, 5*(1), 91–105. <https://doi.org/10.1177/1363460702005001005>
- Attwood, F. (Ed.). (2010). *Porn.Com: Making Sense of Online Pornography*. Peter Lang.
- Aylo: World-class adult entertainment. Some of the internet’s safest platforms. (2023). Retrieved February 11, 2024, from <https://www.aylo.com/>
- Bajhor, H. (2023, April). Whoever Controls Language Models Controls Politics – HANNES BAJOHR. Retrieved September 25, 2023, from <https://hannesbajohr.de/en/2023/04/08/whoever-controls-language-models-controls-politics/>
- Bandelj, N., & Morgan, P. J. (2015, January). Culture and Economy. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)* (pp. 535–541). Elsevier. <https://doi.org/10.1016/B978-0-08-097086-8.10441-6>

- Barthes, R. (1981). *Camera Lucida: Reflections on Photography* (R. Howard, Trans.). Hill; Wang.
- Baudrillard, J. (2001). *Jean Baudrillard: Selected Writings: Second Edition* (M. Poster, Ed.; J. Mourrain, Trans.). Stanford University Press.
- Bell, L., Gladwin, R. S., Hendy, E., Pachnanda, A., Hendy, E., & Dirix, A. (2019, February). Women Tell Us Why They Prefer Porn With No Women. Retrieved February 13, 2024, from <https://www.vice.com/en/article/3kgdwy/women-tell-us-why-they-prefer-porn-with-no-women>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Benjamin, R. (2019, June). *Race After Technology: Abolitionist Tools for the New Jim Code* (1st edition). Polity.
- Bird, A. (2022). Thomas Kuhn. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2022). Metaphysics Research Lab, Stanford University. Retrieved March 25, 2024, from <https://plato.stanford.edu/archives/spr2022/entries/thomas-kuhn/>
- Bitsikokos, L. (2024, April). Living in the Hub: A Platform Study of Desire Semantics - Replication material. Retrieved April 1, 2024, from <https://github.com/lbitsiko/pornhub-categorization>
- Bitsikokos, L., Borlase, K., & Helbing, J. (2023). Tracking Sexism in Pornhub Comments. <https://github.com/mac30123-s23/final-project-pornhub-scholars>
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Bloor, D. (1991, September). *Knowledge and Social Imagery* (2. edition, Ed.). University of Chicago Press.

- Bogost, I., & Montfort, N. (2009). Platform Studies: Frequently Questioned Answers. Retrieved March 17, 2024, from <https://escholarship.org/uc/item/01r0k9br>
- Bowker, G. C., & Star, S. L. (2000, August). *Sorting Things Out: Classification and Its Consequences* (n edition). The MIT Press.
- Burke, K. (2023, April). *The Pornography Wars: The Past, Present, and Future of America's Obscene Obsession*. Bloomsbury Publishing.
- Cattell, R. B. (1966). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*. https://doi.org/10.1207/s15327906mbr0102_10
- Center, B. R. (2023, August). Porn Industry Revenue - Numbers & Stats. Retrieved October 30, 2023, from <https://bedbible.com/porn-industry-revenue-statistics/>
- Chandler, D. (2022). *Semiotics: The Basics* (4th). Routledge.
- Cormen, T. H. (Ed.). (2009). *Introduction to algorithms* (3rd ed). MIT Press.
- Cornyetz, N. (1997, January). Fetishized Blackness: Racial Desire in Contemporary Japanese Narrative and Culture [Section: Literature and Ethnic Discrimination]. In *Literature and Ethnic Discrimination* (pp. 205–232). Brill. https://doi.org/10.1163/9789004656505_016
- Daskalopoulou, A., & Zanette, M. C. (2020). Women's Consumption of Pornography: Pleasure, Contestation, and Empowerment. *Sociology*, *54*(5), 969–986. <https://doi.org/10.1177/0038038520918847>
- De Meo, P., Ferrara, E., Fiumara, G., & Provetti, A. (2011). Generalized Louvain method for community detection in large networks. *2011 11th International Conference on Intelligent Systems Design and Applications*, 88–93. <https://doi.org/10.1109/ISDA.2011.6121636>
- Dey, K., Shrivastava, R., Kaushik, S., & Subramaniam, L. V. (2017). EmTaggeR: A Word Embedding Based Novel Method for Hashtag Recommendation on Twitter. *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, 1025–1032. <https://doi.org/10.1109/ICDMW.2017.145>

- Diakopoulos, N. (2014). Algorithmic Accountability Reporting: On the Investigation of Black Boxes. <https://doi.org/10.7916/D8ZK5TW2>
- D’Orta, M. (1992, January). *Dio Ci Ha Creato Gratis : Il Vangelo Secondo I Bambini Di Arzano*. A. Mondadori.
- Duggan, L., & Hunter, N. (1995, October). *Sex Wars: Sexual Dissent and Political Culture* (1st edition). Routledge.
- Felkins, S. (2015). Fetishizing Southern Brutality: An Intersectional Analysis of Animalistic Dehumanization in Interracial Pornography. *Electronic Theses and Dissertations*. <https://egrove.olemiss.edu/etd/1229>
- Feyerabend, P. (1975). *Against Method* (Fourth Edition). Verso.
- Feyerabend, P. (1987). *Farewell to reason*. Verso.
- Firth, J. R. (1968). *Selected Papers of J.R. Firth 1952-59* (F. R. Palmer, Ed.). Longmans.
- Fleck, L. (1981, August). *Genesis and Development of a Scientific Fact* (F. Bradley & T. J. Trenn, Trans.). University of Chicago Press.
- Foote, J. (2022). A Systems Approach to Studying Online Communities. *Media and Communication*, 10(2), 29–40. <https://doi.org/10.17645/mac.v10i2.5042>
- Fuller, M. (2008). *Software Studies: A Lexicon*. MIT Press.
- Garip, F. (2012). Discovering Diverse Mechanisms of Migration: The Mexico-US Stream 1970-2000. *Population and Development Review*, 38(3), 393–433. Retrieved April 24, 2024, from <https://www.jstor.org/stable/41857399>
- Garlick, S. (2011). A New Sexual Revolution? Critical Theory, Pornography, and the Internet. *Canadian Review of Sociology/Revue canadienne de sociologie*, 48(3), 221–239. <https://doi.org/10.1111/j.1755-618X.2011.01264.x>
- Geertz, C. (2017, August). *The Interpretation of Cultures* (3rd edition). Basic Books.
- Geiger, R. S. (2014). Bots, bespoke, code and the materiality of software platforms. *Information, Communication & Society*, 17(3), 342–356. <https://doi.org/10.1080/1369118X.2013.873069>

- Gentile, K. J. (2010, November). *Sexing the Look in Popular Visual Culture* (New edition). Cambridge Scholars Publishing.
- Glaser, V. L., Pollock, N., & D'Adderio, L. (2021). The Biography of an Algorithm: Performing algorithmic technologies in organizations. *Organization Theory*, 2(2), 26317877211004609. <https://doi.org/10.1177/26317877211004609>
- Goldsmith, K., Dunkley, C. R., Dang, S. S., & Gorzalka, B. B. (2017). Pornography consumption and its association with sexual concerns and expectations among young men and women. *The Canadian Journal of Human Sexuality*, 26(2), 151–162. <https://doi.org/10.3138/cjhs.262-a2>
- Greenacre, M., Groenen, P. J. F., Hastie, T., D'Enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal component analysis. *Nature Reviews Methods Primers*, 2(1), 1–21. <https://doi.org/10.1038/s43586-022-00184-w>
- Hald, G. M., & Štulhofer, A. (2016). What Types of Pornography Do People Use and Do They Cluster? Assessing Types and Categories of Pornography Consumption in a Large-Scale Online Sample. *The Journal of Sex Research*, 53(7), 849–859. <https://doi.org/10.1080/00224499.2015.1065953>
- Heller, T. C., & Sosna, M. (1986, January). *Reconstructing Individualism: Autonomy, Individuality, and the Self in Western Thought* (D. E. Wellbery, Ed.; First Edition). Stanford Univ Pr.
- Holmes, C. (2016). The Colonial Roots of the Racial Fetishization of Black Women. *Black & Gold*, 2(1). <https://openworks.wooster.edu/blackandgold/vol2/iss1/2>
- How Big Is Porn? (2001). Retrieved October 30, 2023, from <https://www.forbes.com/2001/05/25/0524porn.html>
- Hunt, L. (1997). The Many Bodies of Marie Antoinette: Political pornography and the problem of the feminine in the French Revolution. In *The French Revolution*. Routledge.
- Jones, A. (2015). Sex Work in a Digital Era. *Sociology Compass*, 9(7), 558–570. <https://doi.org/10.1111/soc4.12282>

- Jones, A. (2020, February). *Camming: Money, Power, and Pleasure in the Sex Work Industry*. NYU Press.
- Keilty, P. (2018). Desire by design: Pornography as technology industry. *Porn Studies*, 5(3), 338–342. <https://doi.org/10.1080/23268743.2018.1483208>
- Keilty, P., & Leazer, G. (2014). What porn says to information studies: The affective value of documents, and the body in information behavior. *Proceedings of the American Society for Information Science and Technology*, 51(1), 1–11. <https://doi.org/10.1002/meet.2014.14505101067>
- Kitchin, R. (2017). Thinking critically about and researching algorithms [Publisher: Routledge eprint: <https://doi.org/10.1080/1369118X.2016.1154087>]. *Information, Communication & Society*, 20(1), 14–29. <https://doi.org/10.1080/1369118X.2016.1154087>
- Kordela, A. K. (2023). Marx's Affect (and Its Exploitation). *Cultural Critique*, 120(1), 43–60. Retrieved January 28, 2024, from <https://muse.jhu.edu/pub/23/article/893155>
- Kubicek, K., Beyer, W. J., Weiss, G., Iverson, E., & Kipke, M. D. (2010). In the Dark: Young Men's Stories of Sexual Initiation in the Absence of Relevant Sexual Health Information. *Health Education & Behavior*, 37(2), 243–263. <https://doi.org/10.1177/1090198109339993>
- Kuhn, T. S. (2012, April). *The Structure of Scientific Revolutions: 50th Anniversary Edition* (I. Hacking, Ed.). University of Chicago Press.
- Kushner, S. (2013). The freelance translation machine: Algorithmic culture and the invisible industry. *New Media & Society*, 15(8), 1241–1258. <https://doi.org/10.1177/1461444812469597>
- Lanier, J. (2018, May). *Ten Arguments for Deleting Your Social Media Accounts Right Now* (First Edition). Henry Holt; Co.
- Mackenzie, A. (2013). Programming subjects in the regime of anticipation: Software studies and subjectivity. *Subjectivity*, 6(4), 391–405. <https://doi.org/10.1057/sub.2013.12>

- Maćkiewicz, A., & Ratajczak, W. (1993). Principal components analysis (PCA). *Computers & Geosciences*, 19(3), 303–342. [https://doi.org/10.1016/0098-3004\(93\)90090-R](https://doi.org/10.1016/0098-3004(93)90090-R)
- Macpherson, B. (2020). Baudrillard on Broadway: Bio-musicals, the hyperreal and the cultural politics of simuloquism. *Journal of Interdisciplinary Voice Studies*, 5(1), 43–57. <https://doi.org/10.1386/jivs-00015-1>
- Marriott, D. (1996). Bordering on: The black penis. *Textual Practice*, 10(1), 9–28. <https://doi.org/10.1080/09502369608582237>
- Marx, K. (1975). A Contribution to the Critique of Hegel’s Philosophy of Right: Introduction. In K. Marx & F. Engels (Eds.), *Marx/Engels Collected Works* (pp. 175–187, Vol. 3).
- Mazières, A., Trachman, M., Cointet, J.-P., Coulmont, B., & Prieur, C. (2014). Deep tags: Toward a quantitative analysis of online pornography. *Porn Studies*, 1(1-2), 80–95. <https://doi.org/10.1080/23268743.2014.888214>
- McKee, A. (2005). The objectification of women in mainstream pornographic videos in Australia [Publisher: Taylor & Francis eprint: <https://doi.org/10.1080/00224490509552283>]. *The Journal of Sex Research*, 42(4), 277–290. <https://doi.org/10.1080/00224490509552283>
- McLuhan, M. (1964). *Understanding Media: The Extensions of Man*. McGraw-Hill.
- Meehan, C. (2021). ‘Lesbian porn. Anything girl on girl’: Young women’s understandings of their engagement with lesbian porn. *Porn Studies*, 8(1), 21–38. <https://doi.org/10.1080/23268743.2020.1756389>
- Meehan, C. (2023). ‘When you become addicted to it you start having these unrealistic expectations’: Young people’s perceptions of porn addiction. *Porn Studies*, 0(0), 1–16. <https://doi.org/10.1080/23268743.2023.2267593>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *Proceedings of the International Conference on Learning Representations*. <https://doi.org/10.48550/arXiv.1301.3781>
- MIT-Press. (2024). Software Studies. Retrieved March 17, 2024, from <https://mitpress.mit.edu/series/software-studies/>

- Mitrou, L. (2023). Some Introductory Notes. In L. Mitrou (Ed.), *Can the Algorithm... be Ethical, Just, Transparent, Can it judge and manage*. Crete University Press.
- Miyazaki, S. (2012). Algorhythmics: Understanding Micro-Temporality in Computational Cultures. *Computational Culture*, (2).
- Mokhtari, M., & Malek Ahmadi, H. (2017). Pornography and Hyperreality; A Qualitative Study of Experiences of Iranian Users in Using Porno-graphical Products. *10*(38), 71–99. Retrieved March 30, 2024, from http://www.jsfc.ir/article_50903_en.html
- Morichetta, A., Trevisan, M., & Vassio, L. (2019). Characterizing Web Pornography Consumption from Passive Measurements. In D. Choffnes & M. Barcellos (Eds.), *Passive and Active Measurement* (pp. 304–316). Springer International Publishing.
- Morin, E. (2022). *Reveillons-nous!* Denoel.
- Möbner, N. (2011). Thought styles and paradigms—a comparative study of Ludwik Fleck and Thomas S. Kuhn. *Studies in History and Philosophy of Science Part A*, *42*(2), 362–371. <https://doi.org/10.1016/j.shpsa.2010.12.002>
- Mu, J., Bhat, S., & Viswanath, P. (2018). All-but-the-Top: Simple and Effective Postprocessing for Word Representations. *ICLR 2018*. <https://doi.org/10.48550/arXiv.1702.01417>
- Mulvey, L. (1975). Visual Pleasure and Narrative Cinema. *Screen*, *16*(3), 6–18. <https://doi.org/10.1093/screen/16.3.6>
- Nash, J. C. (2014). Black Anality. *GLQ: A Journal of Lesbian and Gay Studies*, *20*(4), 439–460. <https://doi.org/10.1215/10642684-2721366>
- Naughton, J. (2018). The growth of internet porn tells us more about ourselves than technology. *The Observer*. Retrieved October 30, 2023, from <https://www.theguardian.com/commentisfree/2018/dec/30/internet-porn-says-more-about-ourselves-than-technology>
- Neal, J. (2013). Let’s talk about interracial porn. *The Gay and Lesbian Review Worldwide*, *20*(4), 23–27. Retrieved February 13, 2024, from <https://go.gale.com/ps/i.do?p=>

- AONE&sw=w&issn=15321118&v=2.1&it=r&id=GALE%7CA336178182&sid=googleScholar&linkaccess=abs
- Neville, L. (2015). Male gays in the female gaze: Women who watch m/m pornography [Publisher: Routledge .eprint: <https://doi.org/10.1080/23268743.2015.1052937>]. *Porn Studies*, 2(2-3), 192–207. <https://doi.org/10.1080/23268743.2015.1052937>
- Nieborg, D. B., Duffy, B. E., & Poell, T. (2020). Studying Platforms and Cultural Production: Methods, Institutions, and Practices. *Social Media + Society*, 6(3), 2056305120943273. <https://doi.org/10.1177/2056305120943273>
- Nieborg, D. B., & Poell, T. (2018). The platformization of cultural production: Theorizing the contingent cultural commodity. *New Media & Society*, 20(11), 4275–4292. <https://doi.org/10.1177/1461444818769694>
- O’Toole, L. (1998, January). *Pornocopia: Porn, Sex, Technology and Desire* (Main edition). Serpent’s Tail.
- Pasquale, F. (2016, August). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press.
- Patel, K., & Bhattacharyya, P. (2017, November). Towards Lower Bounds on Number of Dimensions for Word Embeddings. In G. Kondrak & T. Watanabe (Eds.), *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)* (pp. 31–36). Asian Federation of Natural Language Processing. Retrieved February 13, 2024, from <https://aclanthology.org/I17-2006>
- Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. <https://doi.org/10.3115/v1/D14-1162>
- Pinch, T. J., & Bijker, W. E. (1984). The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. *Social Studies of Science*, 14(3), 399–441. Retrieved March 24, 2024, from <https://www.jstor.org/stable/285355>

-
- Pires, I. (2024). A Lotus or a Dragon? - The orientalization and fetishization of Asian women's bodies. *Ciência & Saúde Coletiva*, 29(2). <https://doi.org/10.1590/1413-81232024292.03592023>
- Poell, T. (2020). Three Challenges for Media Studies in the Age of Platforms. *Television & New Media*, 21(6), 650–657. <https://doi.org/10.1177/1527476420918833>
- Pornhub. (2022a). The 2022 Pornhub Year in Review — Pornhub Insights. Retrieved August 10, 2023, from <https://www.pornhub.com/insights/2022-year-in-review>
- Pornhub. (2022b). Crash Course: Categories and Tags — Pornhub Model Blog. Retrieved February 11, 2024, from <https://www.pornhub.com/blog/crash-course-categories-and-tags>
- Pornhub. (2023a). Pornhub Insights – Digging deep into the data. Retrieved August 10, 2023, from <https://www.pornhub.com/insights/>
- Pornhub. (2023b, December). 2023 Year in Review - Pornhub Insights. Retrieved February 11, 2024, from <https://www.pornhub.com/insights/2023-year-in-review>
- Powell, W. W., & DiMaggio, P. J. (2012). *The New Institutionalism in Organizational Analysis*. University of Chicago Press.
- Rama, I., Bainotti, L., Gandini, A., Giorgi, G., Semenzin, S., Agosti, C., Corona, G., & Romano, S. (2023). The platformization of gender and sexual identities: An algorithmic analysis of Pornhub. *Porn Studies*, 10(2), 154–173. <https://doi.org/10.1080/23268743.2022.2066566>
- Raunak, V., Gupta, V., & Metze, F. (2019, August). Effective Dimensionality Reduction for Word Embeddings. In I. Augenstein, S. Gella, S. Ruder, K. Kann, B. Can, J. Welbl, A. Conneau, X. Ren, & M. Rei (Eds.), *Proceedings of the 4th Workshop on Representation Learning for NLP (RepL4NLP-2019)* (pp. 235–243). Association for Computational Linguistics. <https://doi.org/10.18653/v1/W19-4328>
- Salganik, M. J. (2019, August). *Bit by Bit: Social Research in the Digital Age*. Princeton University Press.

- Saunders, R. (2020). *Bodies of Work: The Labour of Sex in the Digital Age*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-49016-4>
- Schuegraf, M., & Tillmann, A. (2017). *Pornografisierung von Gesellschaft. Perspektiven aus Theorie, Empirie und Praxis*. Herbert von Halem Verlag.
- Scott, W. R. (2002, January). *Organizations: Rational, Natural, and Open Systems* (5th edition). Taylor & Francis.
- Sebikova, L. (2018). The Ethical Aspects of Techno-Science, Dromology, and Hyperreality. *Komunikácie - vedecké listy Žilinskej univerzity v Žiline*, 20(1A), 32–37. <https://www.ceeol.com/search/article-detail?id=1121251>
- Segal, L. (1998). Only the Literal: The Contradictions of Anti-pornography Feminism [Publisher: SAGE Publications Ltd]. *Sexualities*, 1(1), 43–62. <https://doi.org/10.1177/136346098001001003>
- Sherman, J. G. (1994). Love Speech: The Social Utility of Pornography. *Stanford Law Review*, 47(4), 661–706. <https://heinonline.org/HOL/P?h=hein.journals/stflr47&i=687>
- Stegeman, H. M., Velthuis, O., Jokubauskaitė, E., & Poell, T. (2023). Hypercategorization and hypersexualization: How webcam platforms organize performers and performances. *Sexualities*. <https://doi.org/10.1177/13634607231170174>
- Sun, C., Miezian, E., Lee, N.-Y., & Shim, J. W. (2015). Korean Men’s Pornography use, Their Interest in Extreme Pornography, and Dyadic Sexual Relationships. *International Journal of Sexual Health*, 27(1), 16–35. <https://doi.org/10.1080/19317611.2014.927048>
- Tan, M. J. P. (2018). Pornography: More "Real" Than Reality Itself. *Humanum Review*, (1). Retrieved March 30, 2024, from <https://humanumreview.com/articles/death-in-pornography-the-killing-the-body-and-the-hyper-real>
- Tassis, S. (2023). Can Algorithms be Ethical? In L. Mitrou (Ed.), *Can the Algorithm... be Ethical, Just, Transparent, Can it judge and manage*. Crete University Press.

- Things Are Looking Up in America's Porn Industry. (2015, January). Retrieved October 30, 2023, from <https://www.nbcnews.com/business/business-news/things-are-looking-americas-porn-industry-n289431>
- Thompson, A. (2019, October). Every Story is an Epstein Story [Section: Posts]. Retrieved September 24, 2023, from <https://components.one/posts/every-story-is-an-epstein-story-stoya-pornhub-step-porn>
- Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the Number of Clusters in a Data Set Via the Gap Statistic. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *63*(2), 411–423. <https://doi.org/10.1111/1467-9868.00293>
- Turkle, S. (2005, September). *The Second Self, Twentieth Anniversary Edition: Computers and the Human Spirit* (20th Anniversary ed. edition). MIT Press.
- Vandenbosch, L., & van Oosten, J. M. F. (2017). The Relationship Between Online Pornography and the Sexual Objectification of Women: The Attenuating Role of Porn Literacy Education. *Journal of Communication*, *67*(6), 1015–1036. <https://doi.org/10.1111/jcom.12341>
- Vytნიორგუ, R. (2024). Twinks, Fairies, and Queens: A Historical Inquiry into Effeminate Gay Bottom Identity. *Journal of Homosexuality*, *71*(7), 1605–1625. <https://doi.org/10.1080/00918369.2023.2186760>
- Watson, B. M. (2021). A finding aid to the pornographic imaginary: Implications of amateur classifications on/by reddit's NSFW411. *Porn Studies*, *8*(2), 201–223. <https://doi.org/10.1080/23268743.2020.1830590>
- Watson, M. A., & Smith, R. D. (2012). Positive Porn: Educational, Medical, and Clinical Uses. *American Journal of Sexuality Education*, *7*(2), 122–145. <https://doi.org/10.1080/15546128.2012.680861>
- Weinberg, M. S., & Williams, C. J. (2010). Men Sexually Interested in Transwomen (MSTW): Gendered Embodiment and the Construction of Sexual Desire [Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/00224490903050568>]. *The Journal of Sex Research*, *47*(4), 374–383. <https://doi.org/10.1080/00224490903050568>

- Williams, L. (Ed.). (2004). *Porn Studies*. Duke University Press.
- Wolfe, S. J., & Roripaugh, L. A. (2006). Feminine Beauty and the Male Gaze in The L-Word. *MP: An International Feminist Online Journal*, 1(4), 1–7.
- Wong, C., Song, Y.-D., & Mahanti, A. (2020). YouTube of porn: Longitudinal measurement, analysis, and characterization of a large porn streaming service. *Social Network Analysis and Mining*, 10(1), 62. <https://doi.org/10.1007/s13278-020-00661-8>
- Woolgar, S. (1991). The Turn to Technology in Social Studies of Science [Publisher: Sage Publications, Inc.]. *Science, Technology, & Human Values*, 16(1), 20–50. Retrieved March 24, 2024, from <https://www.jstor.org/stable/690038>
- Wright, P. J., Tokunaga, R., & Herbenick, D. (2023). But Do Porn Sites Get More Traffic than TikTok, OpenAI, and Zoom? *The Journal of Sex Research*, 60(6), 763–767. <https://doi.org/10.1080/00224499.2023.2220690>
- Yin, Z., & Shen, Y. (2018). On the Dimensionality of Word Embedding. *Advances in Neural Information Processing Systems*, 31. Retrieved February 13, 2024, from <https://proceedings.neurips.cc/paper/2018/hash/b534ba68236ba543ae44b22bd110a1d6-Abstract.html>
- Zizek, S. (1992, September). *Looking Awry: An Introduction to Jacques Lacan through Popular Culture* (Revised ed. edition). The MIT Press.
- Zuboff, S. (2020, March). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (Reprint edition). PublicAffairs.