Comparative Analysis of Adult Video Streaming Services: Characteristics and Workload

Raymond Yu, Callan Christophersen[†], Yo-Der Song[†], and Aniket Mahanti[†]
[†] {cchr158, yson715, amah811}@aucklanduni.ac.nz,
University of Auckland, New Zealand

***Abstract*—With the Internet continuing to produce more Web 2.0 services and online adult content seemingly taking up a large proportion of the Internet traffic, pornography services have embraced the shift to user generated content (UGC). This has allowed for more user engagement, forming the so called Porn 2.0. By investigating the characteristics of UGC porn, we can better understand how users are interacting with these streaming services. However, due to the taboo nature of pornography, there has been little work done in this area. In this paper, we examine three of the most consistently popular Porn 2.0 services: PornHub, xHamster, and YouPorn. Using a video corpus of nearly 3 million videos spanning over 10 years, we find that adult videos are commented and rated much less frequently than they are viewed, videos are significantly shorter than the typical TV show, and closer in duration to that of the typical YouTube video. The views for the videos are skewed towards a few popular videos. Video injection rate was variable across the sites. All three sites have on average far more ratings than comments. Video comments are distributed as per the Pareto principle, where a small number of videos receive vastly more comments than the other videos. Most videos receive a moderate number of tags with some more popular videos receiving higher numbers of tags.**

I. Introduction

The ever-growing popularity of video streaming services such as YouTube means an increasingly large portion of web traffic is dominated by these sites. The ability for users to upload their self-generated video content, commenting on content and sharing, has come to be known as Web 2.0 features and is widely embraced by adult streaming sites [1]. Although adult streaming services have been popular for some time [2], there has been limited effort towards investigating the general characteristics of these services.

The majority of prior work in video streaming characterization has focused mostly on YouTube and other popular mainstream video sharing services. We study the characteristics of adult streaming services and identify the commonalities as well as the distinct variances between these adult streaming services and their mainstream counterparts. In particular, we aim to see if it is possible to glean wider social insights from the metadata collected from the adult streaming services. We see this as an important step in the broader understanding of video streaming services and Web 2.0 services as a whole.

Using active measurements, we collected metadata from three of the largest adult video sharing services: PornHub, xHamster, and YouPorn. YouPorn was launched in 2006, while PornHub and xHamster were founded in 2007. These three sites are among the most popular by site traffic according to Alexa rankings [3] and cater mostly to the English-speaking demographic. PornHub and YouPorn are owned by the same company MindGeek, which operates a number of other popular porn sites as well. This allows for cross comparison between competing porn platforms as well as a comparison within the MindGeek platform.

In total, our crawl contained metadata on over 2.9 million videos uploaded over a 10-year period receiving over 354 billion views. This extensive dataset allows us to identify key similarities and differences between these services on aspects such as the popularity of videos, the level of interaction by users, and the uploading pattern of content.

Our results show that adult videos are commented and rated much less frequently than they are viewed, videos are significantly shorter than the typical TV show, and have similar duration as that of the typical YouTube video. A few popular videos are responsible for most of the views. Videos were injected at a variable rate with some sites initially having higher active uploaders than other sites. The number of videos that were commented outnumbered the number of videos which received a comment. There is positive correlation between the number of views and ratings. Video comments are distributed as per the Pareto principle. Tags ranked by number of video uploads followed a Zipf distribution.

Our paper makes two key contributions. First, to the best of our knowledge, this is the first work to perform a comparative workload characterization of multiple large adult streaming services. Second, we identify common characteristics and differences across these services as well as mainstream streaming sites. These common characteristics can be considered as invariants and utilized to update traffic models. We also discuss implications of our results for service operators and end users.

The remainder of this paper is organized as follows. Section II discusses relevant prior work. Section III describes our data collection methodology. Section IV presents our characterization results. Section V discusses the implications of our work. Section VI concludes the paper.

II. Related work

Prior work on web-based video sharing has primarily focused on mainstream sites such as YouTube. Extensive research has been conducted on examining usage and popularity characteristics of the site [4]–[6]. For instance, it has been found that a typical video on YouTube is around 3 to 5 minutes long, indicating that short videos are the norm for

978-3-903176-17-1 / © 2019 IFIP

video streaming services [5], [6]. Video popularity analysis has in the past found that the Pareto principle applies to total views [4]. Gill et al. [6] have shown that the number of video lifetime views follows a Zipf behavior with cut off. It has also been suggested that recommendation systems and common crawling approaches may skew results towards popular content and weed out the unpopular content [7].

Other video sharing services have also been studied for workload characteristics. Mitra et al. [7] analyzed video sharing services such as Dailymotion, Yahoo! Video, Veoh, and Metacafe. This work identified seven invariants focusing on video popularity distribution, use of social and interactive features, and uploading of new content. They showed that popularity of the videos with respect to views followed the Pareto distribution. They also observed certain differences across the sites such as the proportion of multi-time uploaders were twice as large for Veoh compared to Yahoo!.

There has been limited investigation into the adult streaming space. Tyson et al. [1], [8] performed a large-scale study of YouPorn. In this study, they collected 183,000 videos in multiple crawls. The authors compared this dataset with what was known about YouTube at the time. They found that 36% of the videos were uploaded by YouPorn itself. While YouPorn has fewer videos than YouTube, the average number of views per video is higher.

Tyson et al. [9] in another study focused on the social aspects of Porn 2.0. They collected data from over half a million users of PornHub to analyze if users were actually social on porn sites. This study found that the largest group on PornHub was heterosexual men. The majority of users particularly like to interact with younger females. Users were prone to lying about certain characteristics such as age and gender. The vast majority of the accounts were unverified.

Ahmed et al. [10] conducted a large-scale measurement and analysis of online adult traffic. Different from previous works which relied on website data obtained by crawling, this work collected HTTP logs from a major commercial content delivery network which included several dozen major adult websites and their users from different continents.

Our study extends prior work. To the best of our knowledge, this is the first work to present a comparative analysis of adult video streaming services. Our metadata is much larger than prior work and we identify similarities and differences among the popular adult video streaming services, which has not been previously attempted.

III. Methodology

Online user generated pornography comes in a variety of forms: text based (mainly ASCII porn and erotica novels), image based (mostly thumbnail gallery posts, peer-to-peer, and usenet), and streaming based (mostly webcam). In this study, we focus on workload characteristics of image based, thumbnail gallery posted videos hosted by three prominent Porn 2.0 sites (PornHub, xHamster, and YouPorn). To crawl these sites we developed a crawler which has four phases: (1) Collect all the individual category URLs from each site's categories page. (2) Use each category URL to collect all video URLs under each category without duplication. (3) Download and archive the HTML associated with each video URL. (4) Parse each HTML document with regex expressions to extract the video identifier, the uploader username, the number of views, video duration, number of ratings, number of comments, tag details, and the date the video was uploaded. These values are stored in CSV files.

A python script then analyzes these CSV files and computes the statistics, ignoring any null entries i.e. if a video does not have a recorded value for a particular characteristic, then this video is ignored from the analysis.

xHamster and YouPorn have the same incremental indexing technique for their videos and they uniformly apply this technique to all of their videos. PornHub is different and uses three schemes for their video indexing: one is a 19 digit hexadecimal ID number (e.g. a3e2bc6e3ad0f8772db), a 13 digit hexadecimal ID number prefixed by a 'ph' (e.g. ph582cae26b8712), and a 9 digit decimal ID number (e.g. 866619551). Videos indexed with a3e2bc6e3ad0f8772db seem to be the oldest, followed by 866619551, then ph582cae26b8712. However, this can be difficult to determine in some cases as PornHub does not provide granular time stamps for when its videos were uploaded, unlike xHamster and YouPorn. PornHub uses the largest appropriate unit of time to designate the age of a video. For instance, PornHub will date all videos uploaded more than 12 months ago, but not more than 24 months ago, as being uploaded 1 year ago.

To further complicate matters, the videos on these sites are not indexed sequentially in any of these indexing schemes; newer videos do have higher index values but there are large gaps of seemingly unpredictable size between these indexes. For this reason, we can not simply iterate through the video URLs with incrementally larger index values and expect the crawls to finish within a reasonable amount of time. However, this does indicate that from time to time videos are removed. To allow the wider research community to reproduce and extend our work, the dataset is publicly available at `bit.ly/TMA_2019`.

IV. Results

A. Summary of Datasets

Table I summarizes our dataset. The dataset contains metadata on approximately 2.92 million video clips. More specifically, the dataset contains information on 919,746 PornHub videos, 1,199,997 xHamster videos, and 795,683 YouPorn videos. At the time the crawling was performed, these were all the videos publicly available on each site so there is no need for us to develop a sampling scheme for each site. In total, these videos were viewed 354 billion times over their lifetime.

Broadly across all three sites, it appears that videos are commented and rated much less frequently than they are viewed, videos are significantly shorter than the typical TV show, and closer to that of the typical YouTube video of 3-5 minutes. The popularity skew measures the proportion

TABLE I
SUMMARY OF THE DATASET

Values	PornHub	xHamster	YouPorn
Number of views			
Mean	164,872.04	111,675.20	86,984.24
Median	35,493.00	49,624.00	14,212.00
Standard deviation	667,136.36	228,950.15	389,172.08
Popularity skew	0.68	0.51	0.73
Maximum	130,442,547	37,268,554	41,213,946
Video duration (seconds)			
Mean	808.80	1,117.60	724.53
Median	504.00	764.00	383.00
Standard deviation	1,004.83	1,096.77	160,522.09
Popularity skew	0.37	0.32	0.51
Maximum	95,443	65,048	143,165,547
Number of comments			
Mean	5.24	10.07	2.06
Median	1.00	6.00	0.00
Standard deviation	16.47	15.92	10.86
Popularity skew	0.66	0.43	0.83
Maximum	2,074	2,368	1,158
Number of ratings			
Mean	466.05	1627.29	262.39
Median	117.00	790.00	38.00
Standard deviation	1,859.71	3253.38	1,276.12
Popularity skew	0.67	0.48	0.74
Maximum	253,025	536,820	123,837
Number of tags			
Mean	9.81	8.72	8.57
Median	10.00	7.00	8.00
Standard deviation	5.36	5.44	4.38
Popularity skew	0.17	0.24	0.19
Maximum	20	33	49

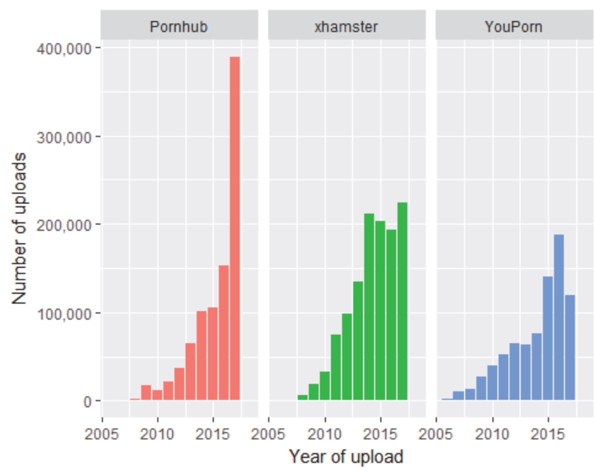

Fig. 1. Number of Video Uploads

of a particular characteristic that the top 10% of videos are responsible for. It shows that for all three sites a small number of videos account for a large fraction of a site's total views.

B. Video Uploads and Uploaders

Video uploads and uploaders is an important characteristic for video streaming websites which measures how frequently users feed content to the system. Figure 1 presents the number of uploads between 2008 to 2017 for the three websites. All 3 websites have shown a growth in uploads. However, the growth in uploads seems to have plateaued in recent years for

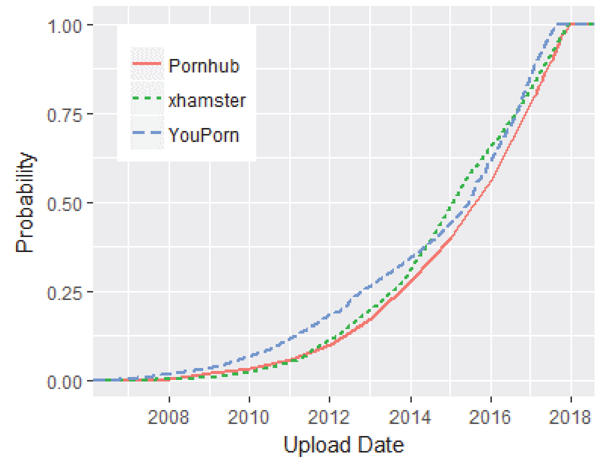

Fig. 2. CDF of Video Uploads per Year

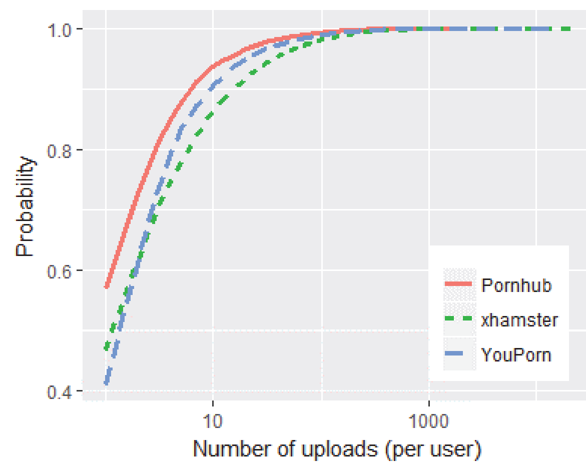

Fig. 3. CDF of Uploads per User

xHamster and YouPorn at about 200,000 videos. PornHub on the other hand continues to grow, almost doubling those of the other two sites. Note that YouPorn's data collection was completed earlier than the two other sites, which could be one reason for this observation. Figure 2 presents the cumulative distribution function (CDF) of the number of video uploaded. We find that across all three sites, roughly 50% of the current corpus were uploaded after 2015.

Figure 3 shows the CDF of the number of uploads per uploader. Here we find that the majority of users across all three sites upload less than 10 videos. We observe that 57% of PornHub users uploaded only one video compared to 47% for xHamster users and 41% for YouPorn users. These rates are similar to what has been found in mainstream streaming sites. PornHub has currently 102,555 active users, whereas xHamster and YouPorn lag behind at 79,656 and 83,225 users, respectively. YouPorn and PornHub being on the same network promote one another on their sites, so we investigated the proportion of users with the same username on both sites. It was found that the two sites shared 6,517 common users,

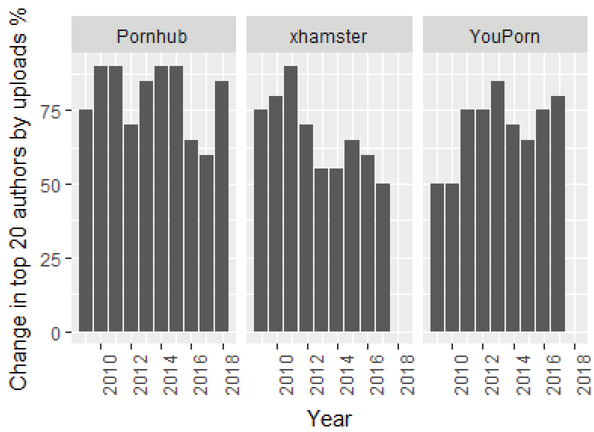

Fig. 4. Change in Uploader Popularity

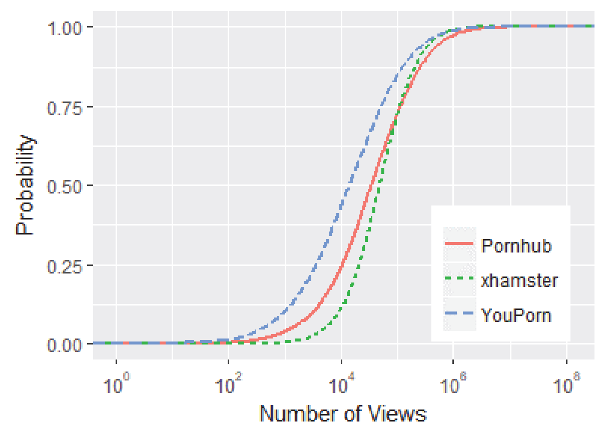

Fig. 5. CDF of Views per Video

roughly 8% of the YouPorn userbase. These usernames had a mean character length of 11 characters, so it was unlikely a large portion of these shared usernames were down to chance alone. What is more likely is the possibility of copycat accounts sharing copied content. These common users are the more active uploaders on their respective sites, accounting for 32% of PornHub views or 53% of YouPorn views. Analysis of xHamster found that it only shared 1,625 users with YouPorn and 1,961 with PornHub.

Figure 4 shows the percentage change in the set of top-20 uploaders based on number of uploaded video between the years. There is a visible significant change in top-20 uploaders. This high churn in top uploaders is expected given the open nature of the industry with low barriers to entry. Various production houses can upload a significant quantity of videos in a short time owing to their large existing catalog of videos, which may contain a lot of content previously released in other places. This can result in large swings in rankings.

We manually analyzed the top-20 uploaders in each data set. For YouPorn and PornHub, most of the top uploaders appeared to be commercial content creators. However, for xHamster, the top contributors appeared to be non-commercial uploaders sharing content acquired from various sources.

C. Number of Views

Views is the predominant measure of popularity for video streaming sites. Across all three sites, PornHub is the most popular with 151 billion total lifetime views as of the time of the crawl. xHamster follows at 134 billion views. YouPorn videos received the lowest number of lifetime views (69 billion), likely due to it having the smallest number of videos in the dataset.

Figure 5 presents the cumulative distribution of number of views for the three sites. Over 90% of the videos across the three sites have between 100 and 100,000 number of views. xHamster has the steepest and sharpest change in gradient of the three sites, and a reasonable amount of overlap between the distributions. Videos with 10 views or less are almost non-

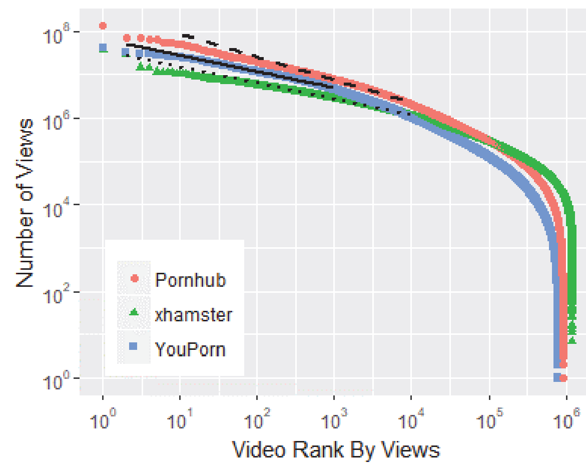

Fig. 6. Video Ranked by Views

existent and we suspect these videos were either deleted or made unavailable.

Power-law behavior is commonly examined for content popularity, often time this behavior holds true for mainstream streaming [6], [7] and so we analyze if it holds for adult streaming. Figure 6 shows the ranked distribution of videos by number of views. We can observe that both PornHub and YouPorn partially follow a Zipf distribution, with the top 20% videos ranked by views accounting for about 83% of total views on the site. xHamster is less skewed by extremely popular videos, showing a slight departure with the top 20% of views accounting for just 70% of views. This may be down to how videos are recommended on the sites. On manual inspection, we found that xHamster shows predominantly videos with only around 100,000 views on its front page. This is a result of xHamster's front page featuring primarily new videos whereas YouPorn and PornHub primarily feature 'popular' videos with the most number of views. The presence of a Zipian distribution for views has implications for the design of caching systems for improving response times and reducing bandwidth consumption. We found that there was an

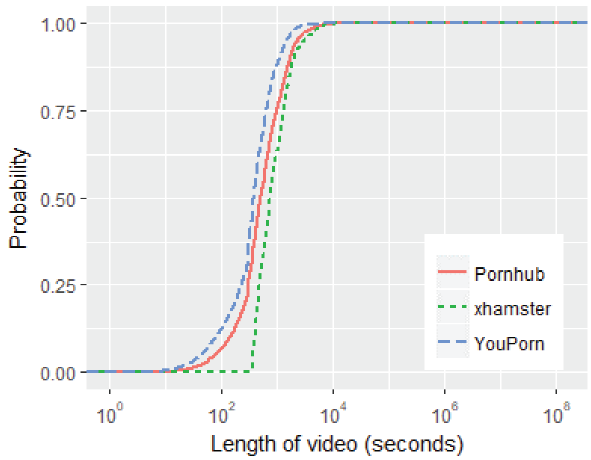

Fig. 7. CDF of Video Duration

apparent Zipfian relationship for the number of views for the top-ranked videos (ranked 10-10,000), while there is a sharp dropoff for the lower ranked videos. The slope values for PornHub, xHamster, and YouPorn were -1.146, -1.148, and -1.314, respectively.

D. Video Duration

Another key characteristic of streaming sites is the duration of its videos. As shown in Figure 7, we observe from the CDF that most videos are of moderate length. We find that 80% of the videos were less than 20 minutes long on PornHub, 25 minutes long on xHamster and 12 minutes long on YouPorn. It is rare to find either extremely long videos on the sites, with only 1% of the videos on PornHub and xHamster exceeding 90 minutes, or very short videos on the sites, with 1% of the videos having a duration less than 30 seconds.

A majority of the outliers that were recorded with a length longer than 24 hours were erroneous duration that changed once the video fully loaded and were in fact only a few minutes long. Others upon manual inspection contained content of only a few minutes with the rest of the duration being a static ad. Similar issues were reported in the duration analysis of mainstream video sharing sites [7].

We further analyzed the video duration by dividing the videos into 1-minute time buckets. We found the largest time bucket was 6-7 minutes for YouPorn and PornHub, and 8-9 minutes for xHamster. The largest bucket contains approximately 12% of all videos for all three sites.

There appears to be a statistically significant difference between the duration of xHamster's videos and the videos of the two other sites. This difference could be due to presence of commercial uploaders on PornHub and YouPorn. Such uploaders tend to upload short previews of longer videos in an attempt to entice users to their paid websites. Having a lower proportion of commercial users on xHamster could be a reason for the longer video duration, with less incentive for amateur uploaders to cut short the duration of their content.

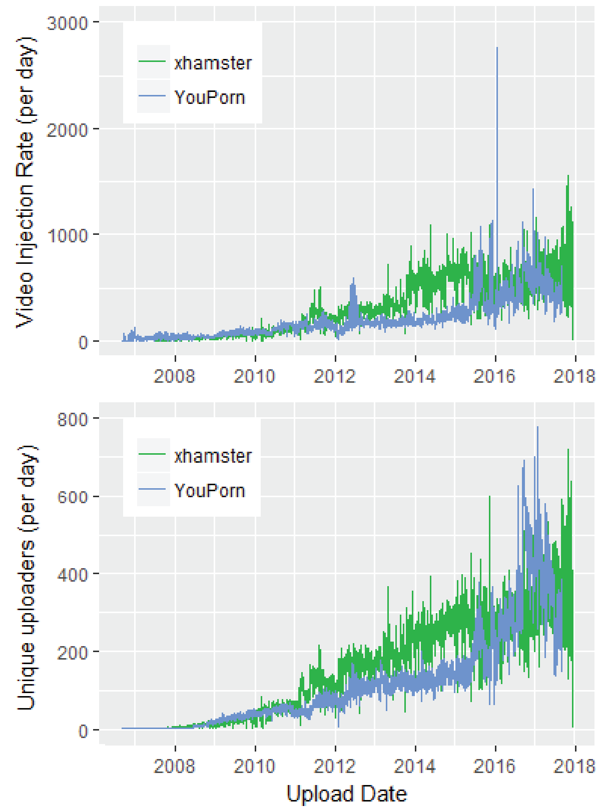

Fig. 8. Timeline of Unique Videos Uploaded per Day (top) and Unique Active Uploaders per Day (bottom)

E. Video Injection Rate

One commonly studied metric of streaming sites is the injection rate of content. A level of churn in content is expected to be necessary to preserve interest in the sites. We compare the unique number of videos uploaded (or the video injection rates) of xHamster and YouPorn. Unfortunately, we are unable to analyze the injection rate for PornHub due to its upload date being rounded. As shown in Figure 8, xHamster has on average 100 more uploaders than YouPorn over the years from 2011 to 2015 before converging to approximately the same rate of roughly 400 unique uploaders a day. Despite this convergence in uploaders there is still noticeably lower number of videos being uploaded to YouPorn. Overall, this greater churn rate seems to contribute to xHamster having higher cumulative views than YouPorn.

F. Ratings and Comments

Ratings and comments are the two common avenues for users to provide feedback relevant to the video content. The rating system used across the three sites are all based on either a like or dislike. We recorded the total number of votes and observe an S-shaped CDF in Figure 9, with less than 10% of videos having no ratings at all. This is in contrast with Yahoo! (55%) and Veoh (43%) which have much higher portions of videos without ratings [7]. All three sites have, on average, far more ratings than comments as expected since it takes much

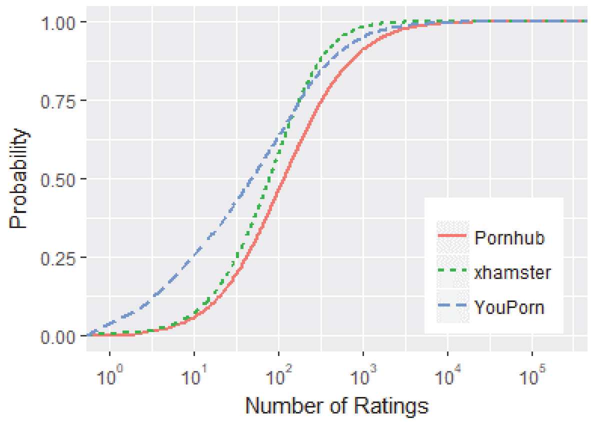

Fig. 9. CDF of Ratings per Video

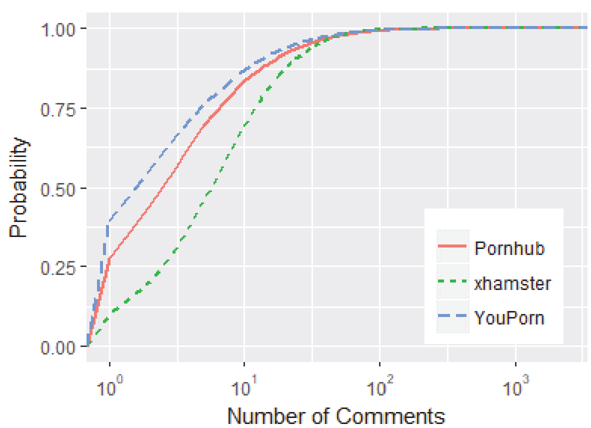

Fig. 10. CDF of Comments per Video

less time and effort for users to interact with videos by rating than commenting. However, the number of ratings is still far less than the number of views the videos receive.

Although most videos have ratings, the number of ratings overall represent a minute fraction of the total views each site receives. The total number of ratings only represent 0.5% of the total views for PornHub and YouPorn and 0.2% of the totl views for xHamster. Despite the low overall use of the rating system, the ratings of a video plays a large part in how content is recommended on the sites. We find the Pearson's correlation coefficient between the number of views and number of ratings is 0.95, 0.80, and 0.82 for PornHub, xHamster, and YouPorn, respectively. This correlation is much higher than what was found for Dailymotion, Yahoo!, or Veoh [7].

The number of comments can provide some indication of the level of interest a particular video has generated and is also a good proxy of the level of engagement viewers have with videos. Figure 10 shows the CDF of the number of comments. We found that 69% of YouPorn's videos have no comments and 26% of PornHub's videos have no comments. This is in contrast with xHamster's 6% of videos without comments. Additionally half of PornHub and YouPorn videos

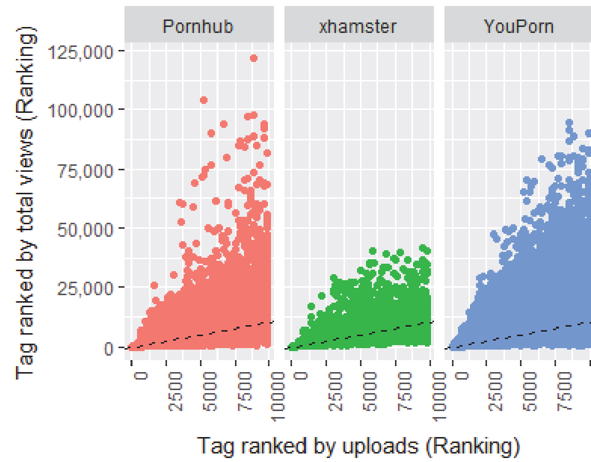

Fig. 11. Scatterplot of Tags Ranked by Uploads versus Views

have less than 5 comments. The low number of views in videos is not the sole explanatory of the low level of comments. In general, the number of comments exhibits some grouping effects with some videos receiving a small number of views and still receiving a number of comments. This is possibly because videos with comments will attract other users to interact and reply. xHamster has more comments per video and more comments distributed across its videos.

G. Supply and Demand of Tags

Video tags can be used to describe a video as well as assist in searching and recommending content to a user. Due to the sites having different ways of adding, changing, and selecting tags, there is a noticeable difference between the sites' distributions for tags. xHamster allows users to add only the tags that correspond to the site's categories. The other two sites allow for more descriptive tags that are outside of the category labels. Since tags are mostly user generated for PornHub and YouPorn, a significant portion of the videos (around 25%) have no or just a single tag. This is not observed in xHamster which does not have freeform tags, but site enforced categories and tags. The total number of tags we observe on YouPorn and PornHub is 2-3 times what is observed for xHamster. Freeform tags may be a reason for the outliers we observe in PornHub and YouPorn, where a large number of tags can be placed, even if the tags are contradictory as there appears to be no restrictions on what can be placed.

We plotted the tags ranked by the number of video uploads versus the tags ranked by the total view count as shown in Figure 11. This plot reveals roughly how good the supply of certain videos is matching its demand or how accurate these tags may be. Since in an optimal system, the most highly used tags should also have the most views and thus follow a linear relationship (as shown by the dotted line in Figure 11). Efficiency of tags seem good for the most popular 100 tags, however significant spread shows after 200 tags. xHamster has a much lower spread than PornHub and YouPorn. This is primarily due to the free form tags used by the sites, which

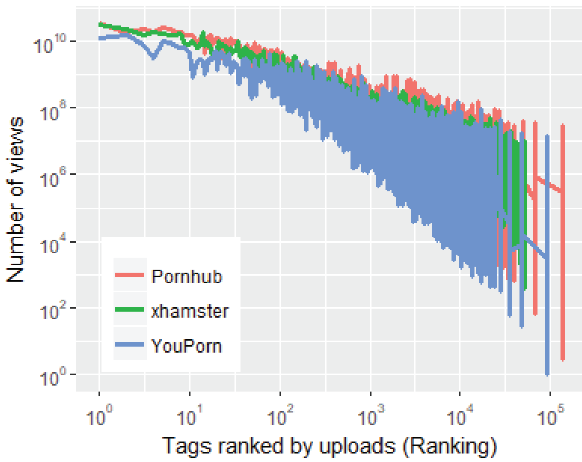

Fig. 12. Rank plot of Tags by Uploads versus Number of Views

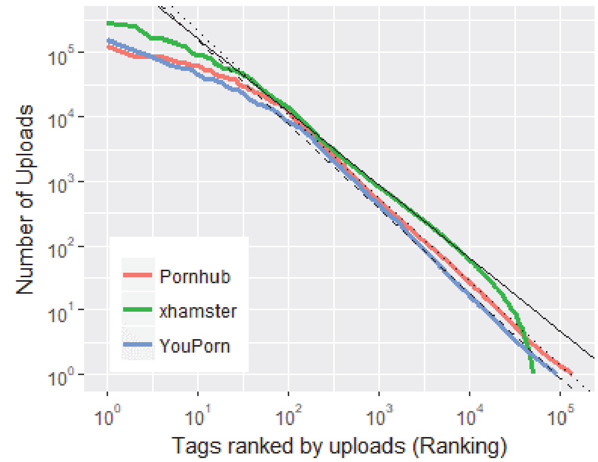

Fig. 14. Rank plot of Tags by Uploads versus Video Count

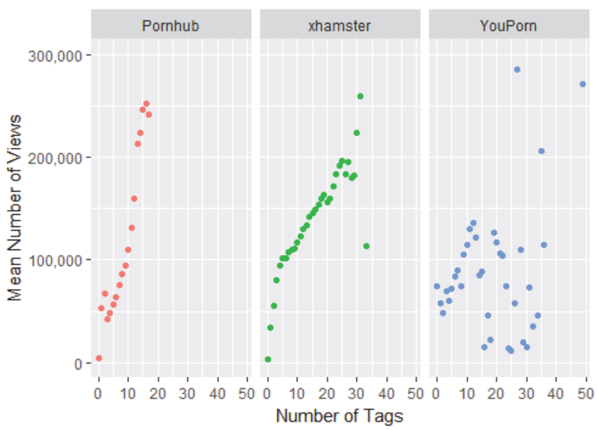

Fig. 13. Scatterplot of Mean Views versus Number of Tags

allows user generated tags. A number of these user created tags are completely nonsensical, such as a string of question marks, dots or random numbers or letters.

When views are plotted against the tags according to tag rank (cf. Figure 12) we find a decreasing pattern following one of a power law, albeit with large amounts of noise. This is particularly random for uncommon tags with a ranking past 100. If it was not for the noise, then there is an apparent Zipfian distribution fitting. This is a likely the result of inefficient tags that do not fully capture what is contained in videos and this is particularly the case for more obscure and abstruse tags.

Figure 13 shows the mean number of views versus number of tags for the three sites. There is a clear increasing linear correlation between the mean number of views and the number of tags for a video on PornHub and xHamster. In particular, we can observe the videos with 0 tags received a very low number of mean views. However, this pattern seems to break down after the 15 tags on YouPorn. There were not enough videos on PornHub with more than 20 tags to compare with xHamster to validate if this is the result of the freeform design of PornHub's tags.

When the number of uploads is plotted against tag rank, a Zipf distribution can be fitted to all 3 empirical distributions (cf. Figure 14). The Zipf distribution is apparent after the top 100 ranked tags. The slope values for PornHub, xHamster, and YouPorn are -1.30, -1.14, and -1.32, respectively. There is a lack of a drop off for PornHub and YouPorn, unlike the power-law relationship for views. Again, we put this down to free form tags and the freedom the sites give to users to add and suggest tags, which means that tags would more closely match the Zipfian distribution behavior found in random word choice. The alpha values of greater than one might suggest that we see these uncommon tags less than we could expect from random words in the English language.

V. Discussion

In this paper, we have given the reader a detailed overview of the interconnected relationship of online UGC video streaming pornography. We summarized the existing literature and found that, whilst others had studied the social habits [9] of porn users and some characteristics of videos on YouPorn [1], [8], there has been limited comparative analysis of adult streaming services. We attempted to fill this gap by performing a cross comparison of video characteristics across multiple porn sites. The characteristics we examined were: uploads, number of views, video duration, injection rate, number of ratings and comments, and tag utilization.

The duration of videos has very similar normalized distributions across all the sites we measured. We therefore suggest that this distribution of duration is a characteristic across all UGC porn sites. The lowest median duration of videos across the three sites we measure was 6 minutes 21 seconds. This is about a minute and a half longer than the average YouTube video duration according to Cheng et al. [5]. This does not necessarily indicate that porn site users watch porn videos for longer than YouTube videos, however it does create the potential for longer viewing times. We leave it to future work to determine if there is a connection between longer duration and longer viewing times. It should also be noted that Cheng et al.'s study was performed over a decade ago. Given this temporal gap in the datasets, it is also possible that the average duration of YouTube videos has increased since that time. This could in turn mean that the average amount of time users view videos has increased across all UGC services.

Both the ratings and views distributions exhibited similar S-shaped patterns. Given the large amounts of overlap between these distributions, we consider both ratings and views to be good candidates for cross UGC porn site characteristics. Their S-shaped CDF distributions combined with the high standard deviations, large ranges, and relatively low averages in Table I indicates a rich-get-richer temporal trend. That is, videos with already relatively large numbers of views and ratings will be promoted by the sites and will therefore be viewed and rated even more. A longitudinal study building on our current dataset would be needed to confirm this assertion.

Our results show that video comments are distributed as per the Pareto principle, where a small number of videos receive vastly more comments than the other videos. From the clear presence of large step sizes for the smaller comment values, we postulate the rich-get-richer principle applies even more so to comments. That is to say, we find it very likely that videos which already have some comments will attract even more comments from others on these sites, due to the social aspects of UGC porn sites. The better a given site is at promoting this social aspect, the more comments its videos will get.

The tagging system is the part of the users' interactions with the videos that these porn sites have the most control over. We believe this is why we do not observe the same amount of overlap in the tag distributions as we do in the distributions of the other characteristics. Even so, all three of these distributions are heavy-tailed distributions and across all the sites we measured, most videos receive a moderate number of tags with some more popular videos receiving higher numbers of tags. Given how dispersed our measured distributions are, we do not believe this is a candidate for a characteristic across UGC porn sites. We remark that our analysis of the utility of tags in the porn sites indicates the importance of category-based browsing [1].

Aside from having a smaller range for most of the characteristics we measure, there are not many remarkable differences in the distribution of these characteristics for xHamster compared to PornHub and YouPorn. This would indicate that MindGeek's ownership of both PornHub and YouPorn does not greatly affect the way in which users interact with these sites when compared to sites not owned by MindGeek. However, it interesting that the average number of comments on xHamster videos is much higher than the average number of comments on the other two sites and the distribution of comments across xHamster videos is more even than for PornHub and YouPorn. We also noted that xHamster videos are on average longer than PornHub or YouPorn videos. This disparity is likely due to the differences in the underlying user population. PornHub and YouPorn have partnerships with commercial content creators who upload video trailers, while xHamster has more amateur uploaders.

VI. Conclusions

In this paper, we examined user generated pornography in three sites, PornHub, xHamster, and YouPorn. We collected the metadata for upload dates, duration, number of tags, number of ratings, number of comments, and the number of views for each video. From this we determined that the distributions for duration, ratings, views, and possibly comments, will hold across other UGC porn sites as well. The number of comments a video receives on average, we believe, has a lot to do with how socially minded the porn site is. If a site promotes social behaviour then more social behaviour, like commentary, will occur. Insofar as the site can control this type of social behaviour, we see the same divergent distributions for videos with relatively few comments as we saw with tags, since each site carries this out in a different way. We described the monopoly MindGeek has over this sort of pornography and how this does not seem to have had an obvious effect over these sites for the characteristics we examined.

The growth in upload frequencies is indicative of the continued and possibly growing popularity of this type of online pornography. As such, we make several suggestions for future work to better understand this growing and dynamic part of the web. We suggest a user study to determine if there is a connection between longer duration and longer viewing times and whether or not users spend, on average, more viewing time per video on porn sites versus mainstream sites. We propound a longitudinal study building on our own dataset, to examine if the user interactions we have measured change substantially over time. Furthermore, other sites such as XNXX, XVideos, LiveJasmin, and RedTube should be studied in a similar manner to discover if the characteristic distributions we uncovered in this work can be found in other porn streaming sites.

References

[1] G. Tyson, Y. Elkhatib, N. Sastry, and S. Uhlig, "Demystifying Porn 2.0: A Look into a Major Adult Video Streaming Website," in *Proc. IMC*, Barcelona, Spain, 2013.

[2] C. Silver, "Pornhub 2017 Year In Review Insights Report Reveals Statistical Proof We Love Porn," *Forbes*, Jan, 2018, [Online: Accessed 4-March-2019].

[3] Alexa, "Alexa Top 500 Global Sites," http://www.alexa.com/topsites, [Online: Accessed 4-March-2019].

[4] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon, "I tube, You tube, Everybody tubes: Analyzing the World's Largest User Generated Content Video System," in *Proc. IMC*, San Diego, USA, 2007.

[5] X. Cheng, C. Dale, and J. Liu, "Statistics and Social Network of Youtube Videos," in *Proc. IWQoS*, Enschede, Netherlands, 2008.

[6] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "Youtube Traffic Characterization: A View from the Edge," in *Proc. IMC*, San Diego, USA, 2007.

[7] S. Mitra, M. Agrawal, A. Yadav, N. Carlsson, D. Eager, and A. Mahanti, "Characterizing Web-based Video Sharing Workloads," *ACM Transactions on the Web (TWEB)*, vol. 5, no. 2, 2011.

[8] G. Tyson, Y. Elkhatib, N. Sastry, and S. Uhlig, "Measurements and Analysis of a Major Adult Video Portal," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 12, no. 2, 2016.

[9] G. Tyson, Y. Elkhatib, N. Sastry, and S. Uhlig, "Are People really Social in Porn 2.0?" in *Proc. ICWSM*, Oxford, UK, 2015.

[10] F. Ahmed, M. Z. Shafiq, and A. X. Liu, "The Internet is for Porn: Measurement and Analysis of Online Adult Traffic," in *Proc. ICDCS*, Nara, Japan, 2016.