A

Measurements and Analysis of a Major Adult Video Portal

Gareth Tyson, Queen Mary University of London
Yehia Elkhatib, Lancaster University
Nishanth Sastry, King's College London
Steve Uhlig, Queen Mary University of London

Today the Internet is a large multimedia delivery infrastructure, with websites such as YouTube appearing at the top of most measurement studies. However, most traffic studies have ignored an important domain: adult multimedia distribution. Whereas, traditionally, such services were provided primarily via bespoke websites, recently these have converged towards what is known as "Porn 2.0". These services allow users to upload, view, rate and comment on videos for free (much like YouTube). Despite their scale, we still lack even a basic understanding of their operation. This paper addresses this gap by performing a large-scale study of one of the most popular Porn 2.0 websites: YouPorn. Our measurements reveal a global delivery infrastructure that we have repeatedly crawled to collect statistics (on 183k videos). We use this data to characterise the corpus, as well as to inspect popularity trends and how they relate to other features, e.g., categories and ratings. To explore our discoveries further, we use a small-scale user study, highlighting key system implications.

1. INTRODUCTION

There have been significant changes in the way the Internet is used. Originally an infrastructure designed for the sharing of computational resources, we now see a diverse range of applications. More recently, the mantra of "content is king" has taken over, giving rise to a large market in content delivery solutions. Video content is particularly popular: by 2019, 80% of all Internet traffic is predicted to be video [Cisco 2015]. This trend has been accompanied by a range of studies into video on demand (VoD) [Yu et al. 2006], Internet TV (IPTV) [Cha et al. 2008; Elkhatib et al. 2014], user-generated content (UGC) [Cha et al. 2009] and catch-up TV [Abrahamsson and Nordmark 2012; Nencioni et al. 2015]. Through these, our knowledge has been expanded and, in many cases, the infrastructures improved.

Despite this, there is one multimedia domain which, even today, is poorly understood: adult video distribution. Traditionally, adult videos were distributed via pay-per-view websites and within peer-to-peer communities (e.g., up to 18% of the files in some regions [Schulze and Mochalski 2009]). Recently, however, increasing amounts of traffic are being generated by emerging YouTube-like websites, colloquially termed "Porn 2.0". These allow users to upload adult videos, so that others can view, rank and comment on them.

Very little is known about Porn 2.0. Despite this, its prominence in the Alexa rankings [Alexa 2014] is undeniable with several Porn 2.0 websites ranking highly. We confirm the scale of adult content in this paper, finding over 111 million requests to a single adult website in just a three day period. Considering this scale, we find it vital to gain a better understanding of the characteristics of Porn 2.0. Whereas this is an important goal in itself, it is also a precursor for devising mechanisms to mitigate its impact on multimedia web infrastructures.

In this paper, we inspect one of the most popular Porn 2.0 websites: YouPorn [YouPorn 2014]. For the last 5 years, it has been amongst the most highly ranked sites listed in Alexa. We began by exploring its delivery infrastructure using DNS and HTTP probes, confirming a significant distribution scale, with servers spread across the globe. Following this, we repeatedly crawled the website to extract information about the videos being created and watched. The corpus consists of 183k videos, with footage spanning in excess of 3 years. Over their lifetime, these videos have collected more than 60 billion views, verifying the website's popularity. In the last 7 days of the traces alone, 912 new videos were added and viewed 38 million times.

Using this data, we characterise the corpus, highlighting how users interact with various aspects of the system. Further, we make a number of observations that provide evidence of its specific properties. Because our data does not provide information about users' personal intentions, we augment our data with a small scale user study to help reflect upon the reasons behind our findings.

Despite being constructed as an adult site comparable to YouTube, the website differs from traditional UGC sites in various ways. First, the number of user-driven content uploads is lower than one would expect. Although amateur uploads are clearly present, we find a notable set of commercial provisions. Second, overall, the number of uploads is far lower than achieved by more traditional UGC services; these videos, however, attain far more views on average. This smaller, more popular content corpus suggests lower operating costs for such websites. We also observe unusual characteristics regarding the way users choose which content to consume. Whilst we verify this independently using a user study, our dataset reveals this in two unexpected ways. First, although adult video content is not expected to age with time, most views are garnered in the first days after upload. Second, the number of category tags a video has and its position in the browsing order has a very important impact: well publicised, easy to find objects collect most views. We demonstrate that the unselective nature of users is a strong driving factor here.

This work extends our earlier work [Tyson et al. 2013]. Here, we dive into the topic of "2.0" by exploring the nature and impact of features such as commenting and rating, as well as extending our investigation of video popularity. We have also introduced a measurement study of YouPorn's infrastructure. A list of contributions and paper roadmap follow:

(1) We offer a large-scale measurement study of adult video distribution. We provide background to the topic (§2), before presenting an analysis of YouPorn's delivery infrastructure (§4). Our datasets (§3) are freely available upon request.
(2) We analyse key characteristics of adult video content (§5), as well as user consumption preferences (§6 and §8). We also investigate the role of other "2.0" aspects in these sites, focussing on rating and commenting (§7).
(3) We explore our findings using a user study, discussing some multimedia system design implications (§9), before providing a comparison between YouPorn and traditional UGC (§10).

Note that our goal is *not* to design new algorithms or tools for this particular multimedia domain. Instead, our focus is to understand and characterise the nature of online adult video distribution. This is something that has been done extensively for other content categories, but has been ignored for this domain. Thus, we believe this is an important precursor to any future steps that can be taken in the area, as well as an important contribution in its own right.

2. BACKGROUND AND RELATED WORK

Pornography is anecdotally the most searched for multimedia content on the web. Cooper [Cooper 1998] attributes this to the web's "triple-A-engine": Accessibility, Affordability and Anonymity. Suler [Suler 2004] expands this into 6 factors, coined as the *online disinhibition effect*. Whereas, much work has gone into looking at who engages in online sexual activities and why [Carroll et al. 2008; Daneback et al. 2012], little is known about the engines that underpin its distribution, especially the expanding Porn 2.0 phenomenon. This has seen websites emerging that, much like YouTube, allow users to upload, view, rate and comment on videos for free.

Some studies have provided estimates of the demand handled by these websites, although few concrete figures are known. One study suggests that certain Porn 2.0 sites (e.g., PornHub, xHamster) can gain up to 16 million views per month. This, however, is extremely low when compared to YouPorn's claim of 100 million page views a day [YouPorn 2012b]. Traffic rates of 800 Gbps have also been estimated by others [Anthony 2012]. Despite this data, we still have quite a rudimentary understanding of the true scale of these services. Regardless, most experts agree that Porn 2.0 is a significant emerging economy that is not, as of yet, well understood [Ogas and Gaddam 2011; Coopersmith 2006; Attwood 2010; Dines 2010]. We confirmed this in our recent work [Tyson et al. 2015], revealing that a non trivial number of people even construct social communities around the topic.

It is therefore surprising to find next to nothing reported on the nature and operation of online adult multimedia services. Instead, various research communities have focussed on specific sub-components, e.g., security issues [Wondracek et al. 2010]; pornographic practices, communi-

Table I. Overview of datasets.

Name	Period	# Vids	# Views
Snapshot	28/Feb/2013	183,639	61 billion
3 Day	3/Mar/2013	183,591	111 million
Daily	1/Mar – 4/May/2013	1656	96 million

ties and subcultures [Attwood 2010]; illegal content dissemination [Hurley et al. 2013], automated recognition and classification [Mehta and Plaza 1997; Hu et al. 2007]; and interest recommendations [Schuhmacher et al. 2013]. To our knowledge, our work presents the first systems-perspective study of an online adult multimedia delivery service. We believe it to be crucial, considering the increasing prominence of video distribution, of which adult media will likely continue to make up a significant proportion in the future [Schulze and Mochalski 2009]. That said, there are several seminal studies into more traditional video streaming systems. These include catch-up TV [Abrahamsson and Nordmark 2012; Nencioni et al. 2015], UGC [Cha et al. 2009; Zink et al. 2009], VoD [Yu et al. 2006] and IPTV [Cha et al. 2008; Hei et al. 2007; Elkhatib et al. 2014]. Studies such as these have provided a range of insights, including content popularity models [Guo et al. 2008], optimised caching techniques [Abrahamsson and Nordmark 2012] and improved delivery schemes [Apostolopoulos et al. 2002]. As of yet, however, we are unaware as to how these principles apply to adult media websites. The rest of this paper therefore explores this topic.

3. YOUPORN DATASET

We have crawled the YouPorn website to obtain information about its corpus and user base. We have selected YouPorn primarily due to its prominence [Alexa 2014]. However, it is also worth noting that the design of most large Porn 2.0 sites are very similar (e.g., in terms of content, layout, search features). Thus, we believe that the results presented here could extend to other similar websites. Our crawls involve downloading the HTML shells of all videos listed in the browsing menus. We extract all metadata associated with each video: the number of views, the number of comments, the upload date, the user who uploaded the video, the video rating, the number of ratings received, and any categorical information. We executed multiple crawls, summarised in Table I.

Our first dataset, which we term the *snapshot* trace, contains information about all videos listed in the browsing menus ($183,639$), collected on the 28/Feb/2013. We observe over 60 billion views of videos with durations collectively spanning in excess of 3 years. To augment this, we also collected a second dataset, which we term the *3 day* trace. To obtain it, we re-crawled the same videos 3 days later (3/Mar/2013). It contained 183,591 videos, 48 having been removed. Using the *3 day* trace, we calculated the evolution of all quantitative metadata.

To build up a finer grained temporal understanding, we also performed smaller scale periodic crawlings to collect a time series of snapshots. The third dataset, which we term *daily*, traced 2172 videos added between the period of 1/Mar/2013 – 4/May/2013. For each video, we collected all metadata on a daily basis. From the full set, we filtered the entries to leave only complete videos in which we had in excess of 21 days recorded. This left 1656 videos, with 96 million collective views. Our data is publicly available upon request.

Before continuing, it is vital to consider the limitations of our dataset. A key limitation is the open nature of YouPorn's metadata. This potentially opens up the system to manipulation, e.g., by video uploaders who wish to artificially inflate their viewing figures. Although this is a possibility, it likely only occurs in the minority of cases due to the resources required. More importantly, our key metric of interest (view counts) is not considered within things like browsing or recommendation ordering, thereby limiting the incentives to manipulate such information. Also, unlike YouTube, we could find no publicly available schemes (e.g., botnets) that would assist in manipulating viewing figures for YouPorn. Lastly, we focus our comparisons with prominent studies that have used the same methodology and assumptions (e.g., YouTube [Cha et al. 2009], Vimeo [Sastry 2012]).

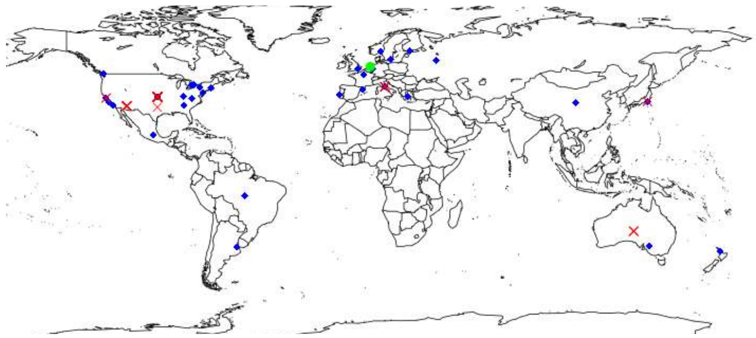

Fig. 1. A map of the delivery infrastructure, measured from our PlanetLab vantage points. Blue diamonds are DNS server locations, red crosses are edge servers and the green circle is the front-end.

Table II. Overview of content delivery domains.

Domain Name	Hosted Content
cdn1.static.phncdn.com	General images (e.g. logo) and supplementary front-end assets (e.g., CSS)
ss.phncdn.com	JavaScript resources
cdn$i$.image.youporn.phncdn.com (where $i = 1, 2, 3...$)	Video images (e.g., thumbnails)
cdn1b.public.youporn.phncdn.com	Videos

4. THE DELIVERY INFRASTRUCTURE

YouPorn operates a content delivery service capable of serving millions of requests a day. We begin by briefly exploring this infrastructure. We probed the website's front-end domain from 56 DNS servers using the PlanetLab infrastructure (Maxmind geo-locatable positions are shown as blue diamonds in Figure 1). These DNS servers were mostly based in Europe and North America, however, we also performed lookups in several other countries including China, Australia, Japan and Brazil. To avoid bias that might emanate from utilising university-based resolvers, we also performed probes using a further 20 third party DNS servers provided via the whatsmydns.net service. Over several days, we sent queries every hour to each DNS server, collecting 60k samples. In all cases, the front-end mapped to a single IP address. Rather than using DNS-based load balancing, our measurements indicate that the website scales by load balancing across servers within a single data centre. Maxmind geo-location shows that this data centre resides in the Netherlands, represented as a green circle in Figure 1. To compliment this, we also fetched the homepage from PlanetLab nodes across the globe; this did not reveal any variations between the content returned.

Following this, we inspected the frontpage HTML. This revealed that all content (videos and images) is served by a set of other domains, whilst the front-end seems only responsible for serving HTML (operating as a central management point). Table II gives an overview of these domains. Interestingly, the domain names in Table II actually belong to PornHub (phncdn.com), another major Porn 2.0 website. By reviewing the HTML of a variety of other prominent Porn 2.0 websites, we found that several of them use the PornHub CDN, including Tube8, XTube and Spankwire. Investigation showed that both YouPorn and PornHub are owned by the same company. It appears that, upon acquisition of these sites, the company has integrated parts of their infrastructure. Arguably, such large parent companies are the only ones with sufficient resources to provide (and monetise) these free adult websites.

To understand the scale of YouPorn's CDN, we also performed DNS lookups using the previously described measurements platform to resolve all of the content domain names in Table II. This collected 173 unique IP addresses, many of which were clustered in the same location and network. In all cases, the domains map to third party providers. The majority of queries (58%) yielded three

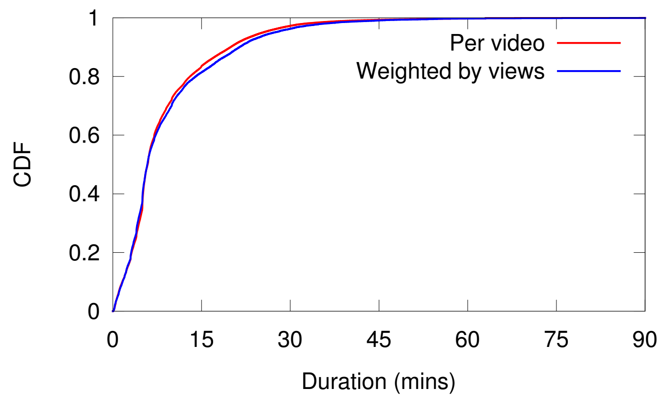

Fig. 2. CDF of content duration.

possible servers for the domains, with the remainder returning one (21%) or two (21%). The geolocation of these servers, where available, is plotted on Figure 1 as blue diamonds. Of course, this is small compared to the scale of a service such as YouTube [Calder et al. 2013]. However, it still constitutes a significant deployment.

5. CHARACTERISING CONTENT CORPUS

In this section we anatomise YouPorn's corpus, focussing on how content is injected and removed.

5.1. Length of Content

We first look at the duration of videos, presented as a cumulative distribution function (CDF) in Figure 2. We observe that most videos are quite short.[1] About 80% are under 15 minutes, with a very small fraction exceeding 45 minutes. The most prominent time range is 5–6 minutes, which constitutes 25% of the corpus. This propensity could have emerged for a number of reasons. An obvious one is the presence of short commercial "teaser" videos (listed alongside amateur contributions) that advertise content from other (pay-based) websites. This often involves providing short previews of videos to entice users to their websites (banners under the videos assist with this). Many other videos also contain just short scenes, without the preambles seen in other media types. The corpus therefore looks very much like a "pick and mix" repository where users can just select snippets of videos that suit their interests. With such time limitations, uploaders (particularly commercial ones) must ensure that only the most interesting elements of their films are seen by viewers. Figure 2 also shows the duration of videos weighted by the number of times they are watched. The curves are near identical, indicating that users do not have a particular preference for one duration but, instead, watch various durations equally often.

5.2. Content Uploads

An important component of Porn 2.0 is the uploading of content by users. We next inspect the frequency at which videos are injected into YouPorn. Figure 3 provides a CDF of the number of daily uploads. An average of 78 videos are added each day, a remarkably low figure when placed in comparison to YouTube [YouTube 2015] and similar sites. For instance, YouTube reported in 2008 to have well over 140 million videos, compared to YouPorn's corpus of 183,639 videos. Despite this, YouTube gets just 100 times more pageviews than YouPorn does [Alexa 2014]. Nevertheless,

[1] Some videos had bogus length fields (e.g., 1000 hours). Consequently, we removed all entries above 3 hours (74 videos), leaving 99.99% of the videos.

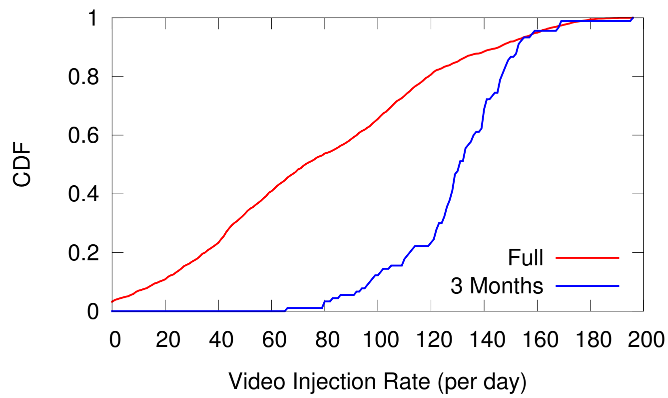

Fig. 3. CDF of daily content uploads.

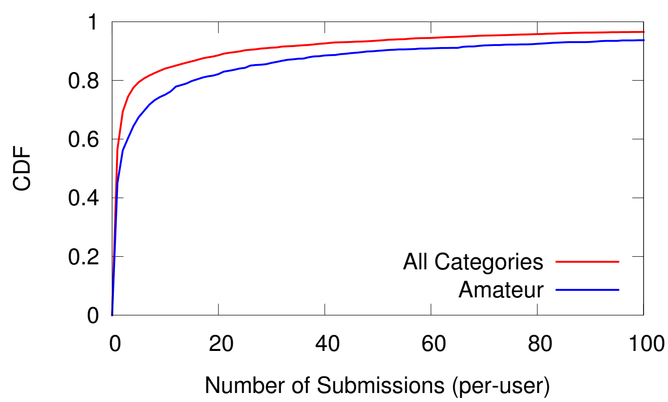

Fig. 4. CDF of per user video uploads.

YouPorn's corpus growth has been accelerating over time, as illustrated in Figure 3, with an average rate of about 140 over the last 3 months of the trace.

We also inspect *who* uploads content. Amazingly, we find only $5,849$ registered usernames that have uploaded content. Note that this includes amateur "Unknown" users with anonymous uploads (33k videos); unfortunately, we cannot discern the upload trends of users within this anonymous set. However, as shown in Figure 4, most registered users (56%) upload just one video, with the bulk (80%) uploading a maximum of 5. Furthermore, Figure 4 highlights the number of per user submissions in the "amateur" category, which intuitively would seem to have more proactive users. Even this category, however, shows very few uploads. Overall, we observe 75 days with no content uploads whatsoever.

To better understand the nature of these uploads, manual inspection was performed. This revealed that many heavy uploaders were actually commercial producers. When excluding the anonymous uploads, the Top 100 uploaders are all commercial (including in the "amateur" category). This trend is best highlighted by one uploader in particular. Figure 5 presents the daily upload rates over YouPorn's lifetime. It shows that 36% of uploads are done by YouPorn itself; a process that started almost 2 years after YouPorn's inception, with an average of 39 uploads per day. These are professional videos that are typically produced by a production studio. We conjecture that this may have been initiated, in part, to ensure a sufficient number of daily uploads. An alternative possibility might be an increasing drive towards monetisation, with YouPorn establishing content partnership agreements with external studios. Without these contributions, the overall average daily upload rates would drop from 78 to 50 video per day. This can be seen in Figure 5 with sustained upload rates boosting the overall uploads after year 2.

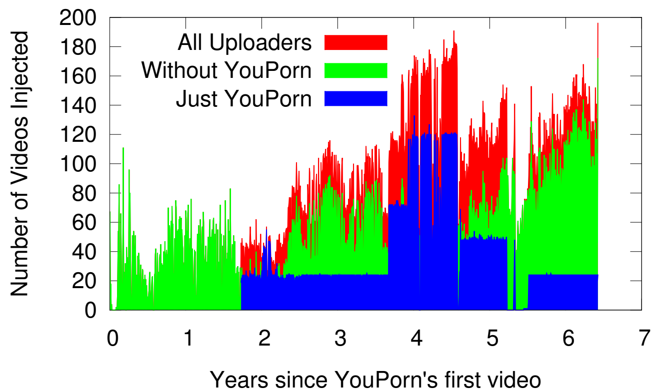

Fig. 5. Daily injection rates (both with and without YouPorn's own contributions): The first two years featured some days with zero new uploads. Beyond year 2, uploads have been significantly boosted by YouPorn itself.

5.3. Content Removal

Of equal importance is the rate at which videos are removed from the corpus. The YouPorn corpus items are identified using an incremental sequence number. We therefore inspect videos in the identifier range of 7692093 – 8300674 (uploaded between Mar/2012 and Mar/2013). The number of removals witnessed in each block is presented in Figure 6; on average, 11.7% of the content is removed. A number of notable spikes can be seen; for example, all videos were removed from a specific 1k identifier block. By excluding these peaks (i.e., any blocks with $> 90\%$ removals), the average percentage of removals comes down to 9.3%. Manual inspection of these removed video's titles shows production studios, indicating potential copyright issues.

From this, one might assume that the majority of videos are active in the repository. However, we did find other video statuses beyond "active" and "removed" (revealed in the HTML response). Most notably, there are a significant number of videos that are classified as being "processed".[2] This status is allocated to a video during the initial stages of its life (e.g., during encoding). It is therefore curious to see that an average of 61% of videos are being processed, sometimes months after being uploaded. We find that a vetting procedure takes place. This is potentially manual, as only a minority of uploads are actually accepted for publication: in the identifier range inspected, only 18% of videos are active. This also offers some explanation as to why injection rates are lower than would be expected.

6. CONTENT POPULARITY

Next, we seek to understand how users interact with YouPorn's corpus. Particularly, relating to the popularity of individual objects, as measured by the number of views they receive.[3]

[2]We also note 8.6% of videos in other states, i.e., "failed" or "not available".
[3]For the remainder of this paper, we use the term *popularity* in reference to a video's accumulated view count.

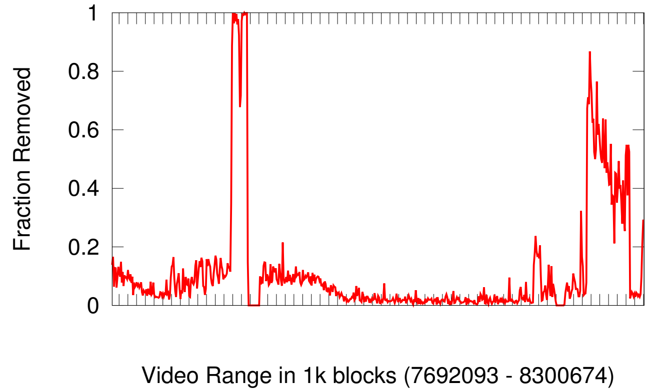

Fig. 6. Removal rate of content across YouPorn's identifier space.

6.1. Measuring the Popularity Distribution

We begin by looking at the number of views (i.e., popularity) accumulated for all videos taken within our traces. We use two different time windows: (*i*) the entirety of YouPorn's lifetime (cumulative) and (*ii*) a three day period. The latter provides us with a recent and short-term perspective irrespective of when the video was uploaded, whereas the former sheds insight into long-term trends.

Figure 7 presents the number of views per video on a log-log plot for both time windows (ordered by rank). Two related observations can be made when inspecting the distributions in Figure 7. First, popularity towards the "head" of the corpus is less pronounced when compared to that of other UGC corpora. In other words, the most popular videos receive far fewer views than would be expected: the top 10% of videos garner only 65% of the total views, whilst this value is 80% for YouTube [Cha et al. 2009] and 82% for Vimeo [Sastry 2012]. Second, the "tail" (or unpopular part of the corpus) gains a relatively larger proportion of the views. From the *snapshot* trace, we find that about 93% of the videos receive at least 10k views. We also observe in the *3 day* trace that 54% of videos receive at least 100 views. In sharp contrast, only 1.9% of Vimeo videos generate more than 10k views. Furthermore, YouPorn's least popular item still manages to receive 226 views, whilst around 53% of the Vimeo corpus even fails to attract more than 200 views. These findings suggest that views are more evenly spread between videos when compared to that of other UGC corpora. This is particularly the case when looking at individual days, where videos uploaded tend to accumulate similar view counts.

The above differences could be explained by two possibilities. One is that YouPorn's videos are generally of a higher quality than in other UGC sites; manual inspection does in fact reveal notable professional content. As such, a higher quality might encourage users to view a more diverse sets of videos. A second explanation is that users have more flexibility in their content selection requirements, i.e., users are not particularly selective in what they choose to watch, thereby resulting in views being more evenly distributed. This could be driven by the smaller corpus, enabling users to more easily browse larger sections of the repository. Our user study (§ 9) indicates that the latter may well be true, with many users having looser interest constraints than traditionally understood. For example, in mainstream VoD services, users often have relatively tight constraints on what they wish to watch. This might be a certain programme, a serialised TV show, or a particular genre [Nencioni et al. 2015]. Without these constraints, however, much larger sets of objects become acceptable for consumption, leading to a lesser skew in the popularity.

That said, the above does not explain the existence of a notable skew, with some objects achieving much higher view counts than others. A key reason is the way viewers discover content to watch. For example, users of other UGC sites may have particular videos in mind, driven by URL links from other websites, e.g., social networks. In fact, it is believed that up to 45% of requests to YouTube

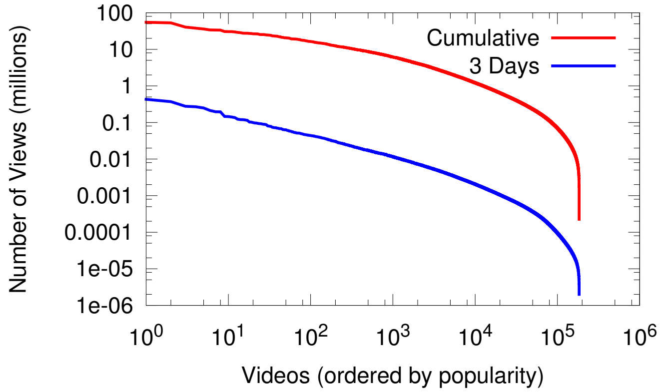

Fig. 7. Number of views per video.

come from social sources [Wattenhofer et al. 2010]. Similarly, Borghol *et al.* [Borghol et al. 2012] found that an uploader's social network is one of the strongest predictors of video popularity. In contrast, users are less likely to share YouPorn links on social networks such as Facebook, or even discuss specific videos with friends (no users in our study had ever shared links). We conjecture that a lack of external referrals helps create an information bottleneck that prevents users from discovering the URLs corresponding to individual videos, thereby forcing viewers to discover videos through YouPorn's built-in facilities (browsing or search). When combined with the inherent user flexibility, this predisposes any "generic" user to retrieve content from the easiest source possible, e.g., front-page listings. To help verify the above assertions, we correlate the number of views per video with the default front page browsing order. Figure 8 presents the results for the *3 day* trace. It can be seen that the majority of views do, indeed, come from easy access items. On average, videos on the front page[4] achieve 55k views, compared to an average of 9k for the top 30 pages. These can then be both contrasted with the overall average of just 603 views per video.

These observations reflect well the type of behaviour one would expect from such a repository. With a corpus in which it is difficult to differentiate objects, only the most dedicated viewers (e.g., ones with special interests) would take the effort to find specific items of interest. More flexible viewers just seek easy access content, which, of course, creates a skew because all users are presented with the *same* easy access content items (note that YouPorn does not adapt its homepage for different regions). However, due to content churn, these objects are quickly pushed from the front pages, flattening popularity into that shown in Figure 7.

6.2. Measuring Temporal Popularity Trends

To gain insight into how views accumulate over time, we look at the distribution of views based on a content item's age. Figure 9 presents a log-log plot of the number of views per video, ranked by view count. Each curve shows the distribution of videos with a specific age. Curiously, videos uploaded a long time ago have not received particularly more views than recently uploaded ones. This is in contrast to past studies (e.g., [Cha et al. 2009; Szabo and Huberman 2010]) that show far greater cumulative views for older UGC videos. For example, in YouTube, approximately 80% of the videos watched (per day) are older than a month [Cha et al. 2009]

To explore the reason, Figure 10 presents the average number of views for videos based on their age. During this short period, people show a distinct preference for recent uploads. Content that was uploaded on the same day as the *snapshot* trace had collected, on average, 28k views just 3 days later, in contrast to an average of only 584 views for older videos in the same period. One theory

[4]This is a conservative estimate as we do not include "featured" videos on the front page.

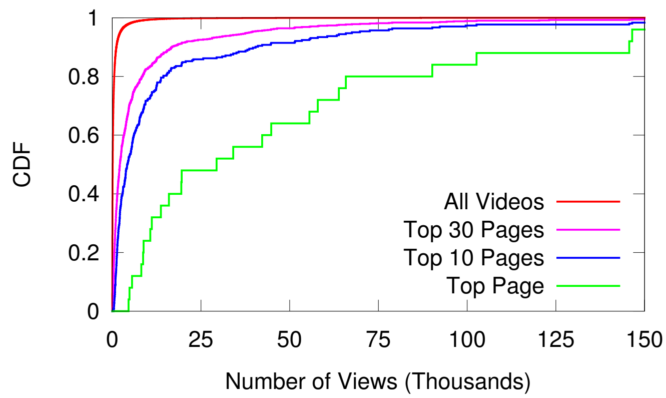

Fig. 8. CDF of number of views per video.

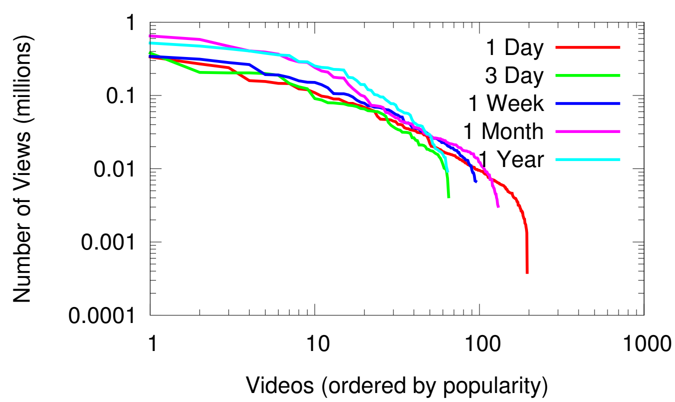

Fig. 9. Number of views per video for different time windows.

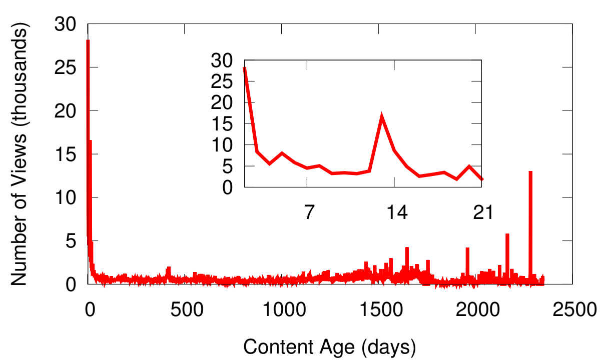

Fig. 10. Number of views per day based on age (inset represents a zoomed version from x=0 to x=21 days).

might be that older videos are simply of a lower quality (e.g., poor resolution), however, we found no evidence for this. Further, Figure 10 does highlight notable exceptions. For example, the third

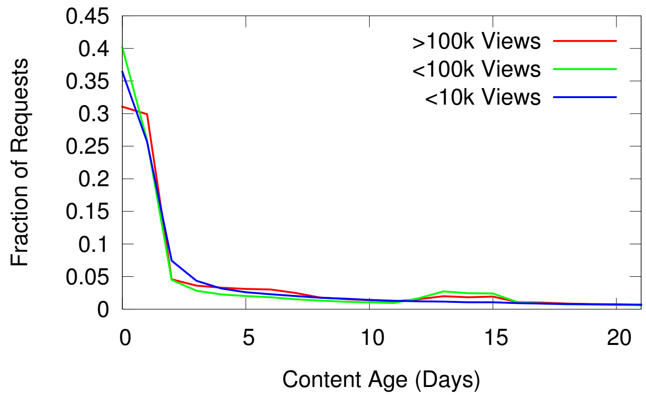

Fig. 11. Evolution of number of views per video over time.

most popular video is actually 6 years old; it therefore seems that older content can be perfectly suitable for viewing.[5]

To understand how the characteristics of videos affect this ageing process, we inspect the *daily* traces to see the evolution of views over a longer period of time. We first separate videos into groups based on the number of views they receive during their first day. We then average the number of views per day received by each video, and normalise that as a fraction of their total view count. Figure 11 shows the outcome. There is a sharp decline in the number of views per day across all popularity groups, with the biggest decrease occurring after the second day. Such trends show why older videos do not exhibit significantly greater popularity than fresh ones: most views are collected in the video's early lifecycle.

Thus, continuing from the discussions in § 6.1, we draw a conclusion. Content in YouPorn's front page is ordered by upload date (and then rating), making recent uploads the easiest to discover. Accessing content older than one day requires browsing through ≈4–5 pages of listings, a process which clearly some users are unwilling to undertake (likely due to the similarity of many videos combined with user flexibility). Only extremely popular videos ($>$ 100k views) can resist this immediate decay, with similar viewing figures being recorded on the first and second day.[6] However, after the third day, even popular videos will be pushed down by ≈10 pages, making it significantly harder for users to locate them. Hence, on the third day, they begin to exhibit the traits of their less popular counterparts. Future views are restricted to those who are prepared to more proactively seek content. Relating this to traditional UGC, Crane and Sornette [Crane and Sornette 2008] provide a classification of video types (memoryless, viral, junk and quality). This would place typical YouPorn videos into the category of junk. These are videos that experience a short burst of activity, followed by a popularity collapse. In contrast, such videos belong to the smallest category in YouTube [Crane and Sornette 2008; Pinto et al. 2013].

7. ANALYSING YOUPORN'S 2.0 FEATURES

The previous section has highlighted clear trends in video popularity. An interesting question is how these trends relate to other aspects of YouPorn. Of particular interest is the use of "Web 2.0" features including video rating and user commenting, which we focus on in this section.

[5]Note that special browsing options make such videos easy to find (e.g., "most viewed").
[6]Highly rated videos will appear at the top of the browsing list for the previous day.

7.1. Video Commenting

Commenting has become a key element of Web 2.0. Figure 12 presents a CDF of the number of comments received by each video. It should be noted that the commenting system was not initially available in YouPorn, it was introduced a few years later. Figure 12 therefore also presents data just from videos uploaded during the last three years (labeled as *recent*).[7] In contrast to other UGC systems [Thelwall et al. 2012], 37% of videos receive no comments whatsoever, with 75% receiving below 5. Clearly, users are not socially engaged in the same way YouTube viewers are. This is perhaps not surprising considering the nature of the content.

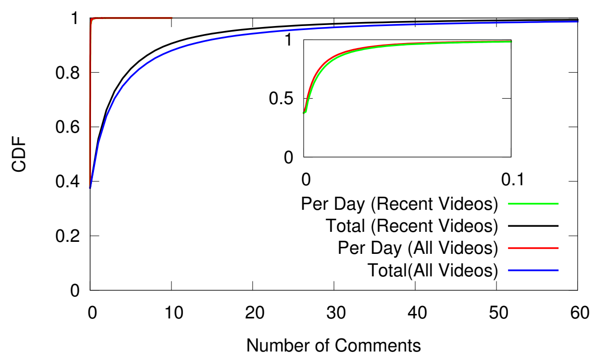

Fig. 12. CDF of number of comments per video.

To normalise the differences in video age, Figure 12 also includes the per day commenting rates taken from the *snapshot* traces. This is calculated by dividing the number of comments by the age of the video. This reveals surprisingly low values. In fact, it shows that 99% of videos receive fewer than 0.1 comments per day, with only a tiny fraction gaining multiple daily comments. To explore the reasons for this, we inspect the temporal trends of commenting. Figure 13 presents the average number of comments received per video over time, taken from the *daily* traces. For completeness, the number of ratings is also shown. Whereas the distribution of ratings follows a very similar trend to that of viewing figures, commenting actually peaks on the second day after a video's release. Initial comments can be seen to often elicit responses, which increases commenting rates over time [Lange 2007]. That said, these rates rapidly diminish after the second day, along with ratings (and viewing figures as discussed previously). Hence, we witness extremely low per day commenting rates, as older videos generally cease to collect comments soon after creation.

7.2. Video Rating

Another 2.0 feature is the rating system. This has always been available within the YouPorn system; initially, this involved rating videos from 1 to 5, whilst since a "like" and "dislike" approach is taken. In both cases, the values are converted to a percentage figure allowing easy comparison. Figure 14 presents a CDF of the number of ratings per video. Users are far more engaged with the rating system; fewer than 1% of videos receive no ratings whatsoever.

To look into how these ratings emerge, Figure 15 presents a CDF of the ratings given to each video taken from the *daily* traces. The figure plots both the initial rating given and the final rating observed. Interestingly, a variety of initial ratings can be seen, rather than a single standard value. The reason behind this is unknown, however, we conjecture that content partners could perhaps influence the initial rating. This is an extremely powerful tool due to the use of rating-based ordering in the

[7] We include both time windows because since the introduction of the commenting feature, *all* videos can be commented on. Excluding earlier videos would exclude ≈500k comments.

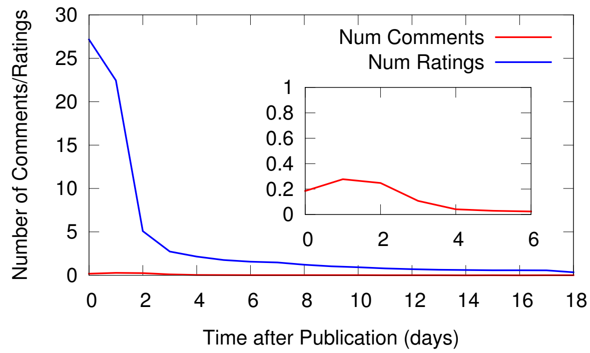

Fig. 13. Average number of ratings and comments per video over time.

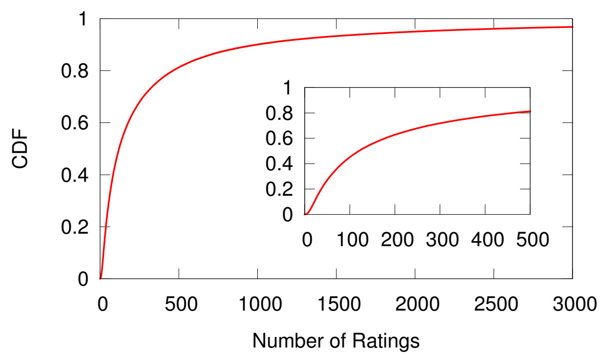

Fig. 14. CDF of number of ratings per video.

browsing options. After a video becomes live, a noticeable shift can also be observed in which ratings, on average, increase after publication. Users are more inclined to “like” than “dislike” a video; in fact, fewer than 5% of videos have a ranking below 50%. This could suggest that either YouPorn possesses a highly appreciated corpus or, alternatively, users only take the effort to rate a video when they enjoy it.

This theory can be inspected by looking at the number of times users choose to rate a video. If the theory is correct, videos with many ratings will also be highly rated. As well as presenting the absolute values, we normalise the number of ratings as a percentage of the views received. Figure 16 presents the results, comparing the number of views against the achieved rating. It can be seen that a strange divergence is present. In absolute terms, videos with many ratings get higher scores (blue triangles), whilst, in normalised form, videos with many ratings actually get lower scores (red crosses). The former would indicate that satisfied users have a greater propensity to rate, whilst the latter would indicate that, actually, dissatisfied users have the greater propensity. In practice, the absolute values, however, are heavily skewed by a few very popular videos — the more views a video receives, the higher the probability that somebody will rate it. This can be seen in Figure 13, where temporal rating trends are near identical to viewing figure trends. Thus, in this circumstance, the number of ratings simply becomes a proxy metric for the number of views. When looking at the normalised version, it then becomes evident that users, in fact, tend to rate videos that they are unhappy with far more than those that they are satisfied with.

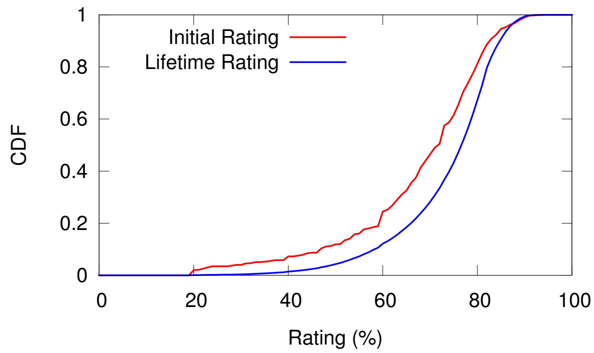

Fig. 15. CDF of ratings per video.

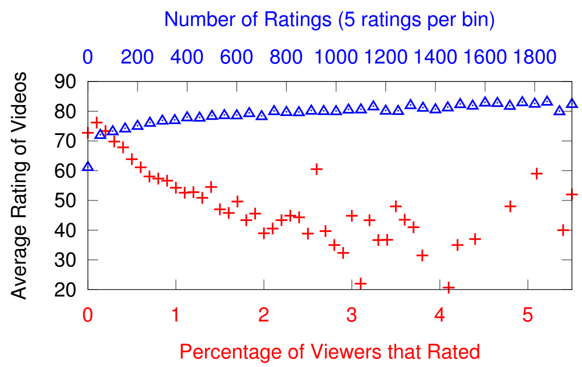

Fig. 16. Comparison of video rating vs. number of ratings.

7.3. An Alternative Popularity Metric?

The above discussion has introduced an alternative measure of the concept of "popularity". In fact, it could be argued that the ratings are actually a superior measure of this concept (as viewing something doesn't necessarily mean it was enjoyed). To explore the relationship between these metrics, Figure 17 presents the average number of views received for videos with a variety of ratings. It can be seen that a relationship does exist, however, it is not linear as one might expect. A very curious trend is exhibited in which videos rated as high as 70% get no real increase in views when compared against videos rated as 30%. This trend continues with only videos rated in excess of 80% showing notable increases. This then increases dramatically when reaching 90%. The peculiar divergence in Figure 16 then begins to make sense. The rating system within YouPorn seems to have suffered from hyper inflation, in which only the "super-rich" can survive. Evidently, it is very difficult for content to differentiate itself in this corpus with only the highest quality and most visible content receiving high viewing figures. In such cases, content with mediocre ratings is left ignored, just like content with very low ratings.

Ratings are therefore a vital component in selecting which videos to watch, offering a powerful insight into potential popularity. This could be seen as viewers specifically targeting well rated content. However, in practice, the nature of YouPorn's browsing system guides users automatically towards these highly rated objects (the ratings are used as a secondary ordering in the browsing menus). As such, it becomes quite difficult for users to gain access to less well rated objects, particularly once they are older than a couple of days. This property forces users towards a small subset of content (an attractive property from a content caching perspective).

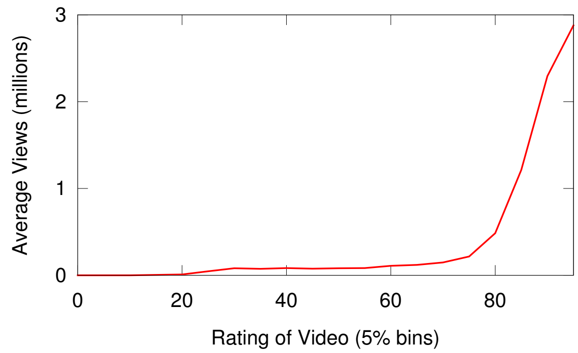

Fig. 17. Comparison of popularity vs. rating.

8. EXPLORATION OF CATEGORIES

So far, the corpus has been inspected as a single collection. Category tags, however, can provide semantic insight into the nature of content by separating the corpus into subcomponents. This allows us to inspect more targeted groups of content.

8.1. Category Overview

A video can be associated with one or more tags that provide a semantic categorisation of content type, e.g., "amateur", "romantic". 62 categories were available at the time of the crawl. After upload, a video can be allocated a category. Later on, further tags can be added by viewers (assuming sufficient numbers). Whilst this community-driven approach means that videos could be incorrectly tagged, the category-specific web pages offer an important portal into content of a broad type. We emphasise that categories are not free-form tags but, instead, limited to certain well defined terms.

Figure 18 depicts the distribution of videos in each category. Due to their explicit nature, the categories have been pseudo-anonymised using their first two letters[8]. A strong skew emerges: the top 20% of genres contain 57% of the videos, whereas the bottom 20% contains under 2%. Only 11% of content belongs to the bottom 50%. This all signifies that YouPorn's categories possess a much longer tail than previously observed in UGC corpora [Cheng et al. 2008; Liikkanen and Salovaara 2015], also with a wider range of possible categories.

Next, we investigate category popularity. We begin by examining the number of views collected by each category, plotted in Figure 19. Populous categories are observed to receive the majority of views, a trend similar to that of Figure 18. An exception is that of "NU" (videos without a category) which does not fare well in proportion to its size: 17% of the corpus is untagged, yet these videos collect only 2% of the views. This confirms that category tags are extremely important in YouPorn.

8.2. Category Efficiency

In a sense, YouPorn could be considered as a form of marketplace where uploaders provide their videos for consumption, competing for views. Thus, ideally, supply for a category would exactly match its demand. This can be captured through the notion of efficiency. An efficient corpus is one in which the fraction of views for a category is equal to the fraction of the corpus that lies in the category. We therefore next inspect how distant each category is from efficiency. The level of *inefficiency*, $\mathcal{I}$, of a category is calculated as follows:

$$\mathcal{I} = \begin{cases} \frac{V}{C} - 1, & \text{if } V > C. \\ -(\frac{C}{V}) + 1, & \text{otherwise.} \end{cases} \tag{1}$$

[8]Mappings at http://www.eecs.qmul.ac.uk/~tysong/yp/mappings.txt

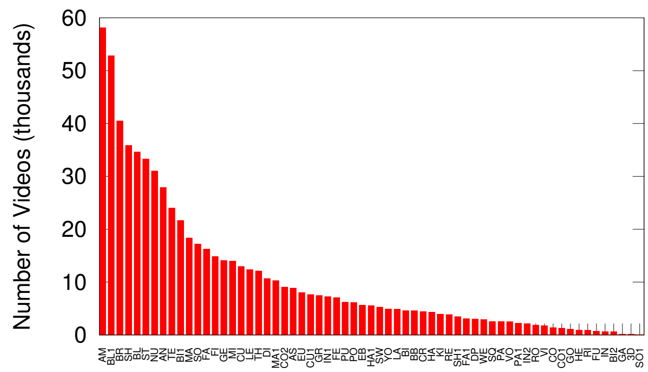

Fig. 18. Number of videos per category (ordered by number of videos in the category).

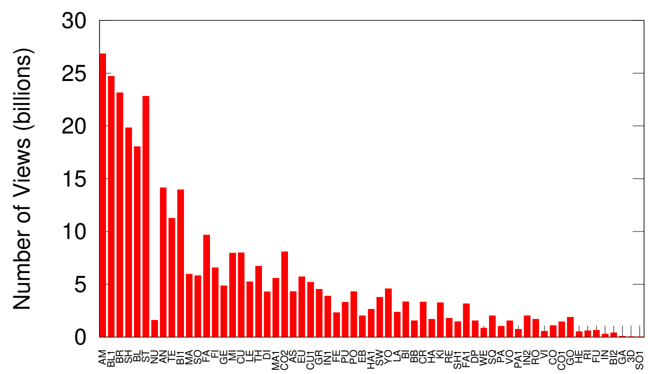

Fig. 19. Number of views per category (ordered by number of videos in the category).

where V is the fraction of views that the category receives, and C is the fraction of the corpus that the category constitutes. A video can be tagged in multiple categories, leading to two ways of calculating V. The first approach attributes one view to each category that the video is tagged with; therefore, a video with two categories, say "ST" and "BR", will have one view allocated to each. A side effect is that the number of views is artificially inflated. Hence, the second approach (termed *weighted inefficiency*) attributes views to categories by splitting the views of a video equally between all κ categories it belongs to. The number of views of the video attributed to that category are then factored by $\frac{1}{\kappa}$. Henceforth, we allocate views counts to each category using the former method, unless otherwise stated.

We can use inefficiency as a normalised measure of category popularity. $\mathcal{I} > 0$ indicates that a category receives a large number of views in proportion to its size. On the other hand, $\mathcal{I} < 0$ signifies that a category collects disproportionately fewer views than expected. More generally, a value above 0 represents a category in which "demand exceeds supply" (popular), and a value below 0 represents a category in which "supply outstrips demand" (unpopular).

The category inefficiency levels are shown in Figure 20. We see that using the absolute number of views a category receives is somewhat misleading. The category with the highest viewing figures ("AM") is actually unpopular in normalised terms ($\mathcal{I} = -0.11$). This category clearly collects views through its size in the corpus rather than through a high demand for the genre. Inefficiency is

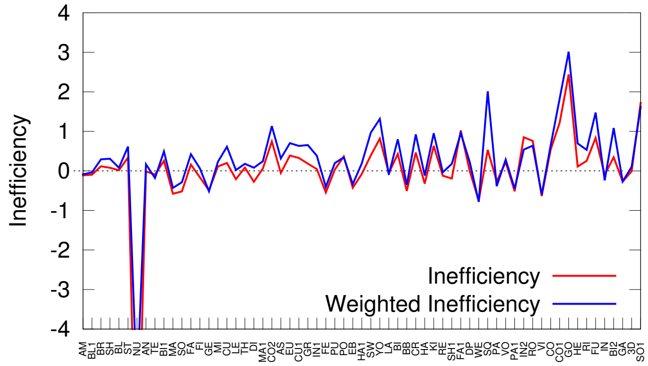

Fig. 20. Inefficiency of categories (ordered by number of videos). For null, $\mathcal{I} = -8.96$ and weighted $\mathcal{I} = -5.51$ (this is cut-off to improve readability).

observable in all other categories too: 27 categories have $\mathcal{I} < 0$, whilst 35 categories have $\mathcal{I} > 0$.[9] "BL" comes the closest to efficient (0.016), alongside 14 others that are between -0.1 and 0.1. In other words, the category make-up of the corpus does not match its consumer-base directly.

8.3. Colocation of Categories

Categories clearly have a critical impact on video consumption. However, we posit that the ability to associate a video with multiple categories (i.e., category colocation) could undermine the aforementioned insights. To examine this, Figure 21 presents the category inefficiency (previously displayed in Figure 20) alongside the average number of colocations per category. The latter shows, for example, that a video tagged with "AM" is also tagged with an average of 6 other categories. There appears to be a strong correlation between normalised popularity and colocation, with a correlation coefficient of 0.66. For example, the "GO" category gets 3 times as many views as could be expected from its proportion of the corpus. It is, however, the most colocated category: "GO" videos are placed in 10.36 categories on average. 4 out of the top 10 categories, ranked by colocation, are also in the top 10 ranked by normalised popularity. Clearly, the high exposure afforded by category-based browsing menu is very beneficial for object popularity.

To validate the impact of colocation, Figure 22 compares the mean number of views against the number of categories a video is part of. There is a near linear trend in which videos belonging to multiple categories get more views. Videos lacking a category receive just 51k views on average, compared to 452k for those with 5 categories: category tags are a significant factor that contributes to views. It is even possible that some uploaders exploit this observation to improve their view counts.

9. DISCUSSION OF USER CONTENT CONSUMPTION

This paper has explored a number of observations that could be taken as indicating that users do not visit YouPorn with specific videos in mind: duration does not matter (Figure 2); users tend to simply go for the most easily accessible videos (Figure 8); the number of views heavily depends on front page listings (Figure 11) and the number of tagged categories (Figure 22). These lead us to argue that the consumption patterns of many users are rather flexible. This, however, is difficult to discern solely via our prior datasets as they cannot tell us individual user preference. Thus, to explore this elasticity, we expand our data by performing an anonymous user study involving 47 respondents

[9] Interestingly, many of the more unusual niche categories (e.g., "GO", "IN2", "FA1") are popular in normalised terms.

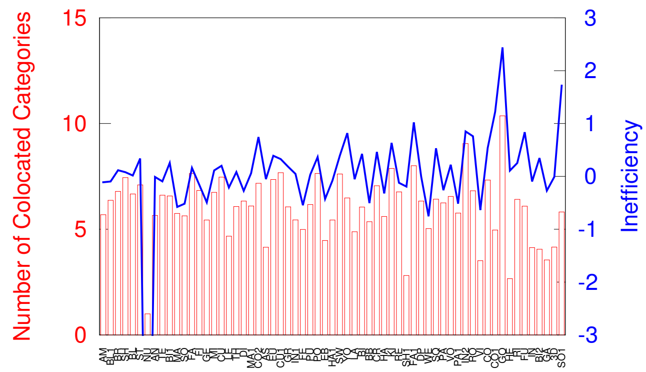

Fig. 21. Average colocation of videos per category (ordered by number of videos in the category).

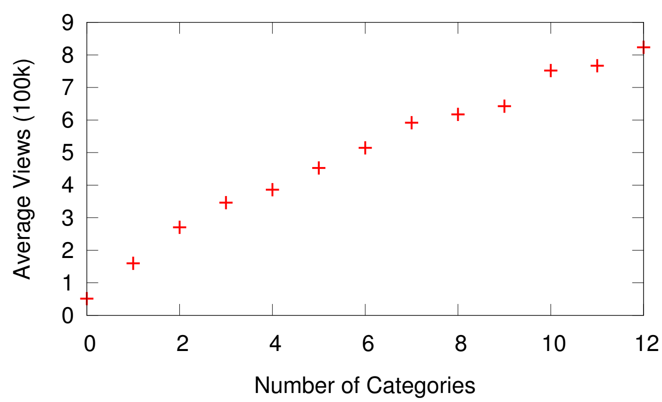

Fig. 22. Number of views vs. the number of categories allocated to video.

recruited via online social networks and mailing lists. This allows us to better explore our earlier conclusions through the lens of end user preferences and intents.

Conforming to our suspicions, we find that 85% of survey participants find it easy or just slightly difficult to find content of interest, with the remainder indicating that they found it difficult. 43% of survey participants also said that over 6 out of every 10 videos they found match their interests, whilst only 30% said that under 4 out of every 10 videos satisfy them. In other words, a typical user can be satisfied by a wide variety of potential videos.

Such flexibility sheds light on some of the aforementioned results. Chiefly, it confirms our expectations that consumption choices are largely guided by the order in which content is displayed on the default front page and on category-specific pages: when asked how they find content (multiple answers were allowed), the survey showed that only 21% of users ever visit the site with a predetermined video in mind. A large share of the respondents instead use “flexible” ways to discover content: 62% rely on browsing the front page, 60% use category-based browsing, and 51% utilise search features. This is unlike many other content repositories where users visit mainly with predetermined videos to consume, e.g., sports [Brampton et al. 2009] or prime time shows [Elkhatib et al. 2014]. Further, only 9% visit YouPorn directly through links from other sites, whilst no user had ever shared links via social networks. This is in contrast to YouTube users, for example, who rely more on web search engines [Cunningham and Nichols 2008] and external links [Wattenhofer et al. 2010], rather than browsing.

The above observations indicate that some flexible users could be easily satisfied with a small set of shared videos taken from a large range of acceptable ones. With many users listing more than 60% of content as acceptable, this seems highly feasible. We therefore argue that this flexibility should be leveraged by the system. Specifically, where many videos could satisfy the user, the content distribution infrastructure could guide people towards those that have a low network cost (e.g., available nearby). This would improve user quality of experience, reduce overheads and free up distributed storage space for other more niche content. Such guidance could easily be implemented by promoting low cost objects at the top of browsing menus, e.g., place nearby videos that match user interests on the front page.

Such novel approaches may be necessary as our survey indicates that previous optimisations such as prefix caching would not work. Only 11% of respondents said they watch videos linearly, with the majority using skip features to find parts of interest. Caching video prefixes (early parts of the video) would therefore be ineffective. Further, as 40% of users state that they normally load multiple videos in parallel, the load on the server would be much larger per-user than in other video domains. A particularly unusual property of this domain is the propensity to “switch” between videos: 15% of respondents said they tended to load videos in parallel and switch between them.

The survey also suggests that only a small amount of (generic) content would be needed to satisfy the demands of many viewers: 86% of users watch under 10 videos a session, whilst 43% watch 3 or less. That said, delivery optimisation should not be done at the expense of a user’s quality of experience. A key challenge would be to ensure that this small body of content is kept sufficiently “fresh”. Only 23% of survey respondents said that they would watch a video multiple times. Thus, it is important that users are constantly provided with novel items.

10. COMPARISON WITH TRADITIONAL USER GENERATED CONTENT

A number of issues have been discussed in the previous sections pertaining to the characteristics of YouPorn. A natural point of comparison is that of “traditional” UGC platforms, such as YouTube. Here, we summarise these comparisons. Table III delineates the key points, based on various related studies of YouTube. Where appropriate, we also include measurements of Vimeo. The key differences derive from the content type on offer. YouTube is a generic platform that hosts a wide variety of material, whilst YouPorn (and equivalent services) are highly targeted. That said, it is worth noting that even targeted adult websites can contain quite diverse content, with YouPorn possessing 62 active categories at the time of the crawl (far more than YouTube’s 15). This means that YouPorn has a far longer tail of categories compared to YouTube, where “Music” is the most popular by views [Liikkanen and Salovaara 2015].

An obvious point of similarity is that both providers encourage third parties (notably amateurs) to upload their own content. An important difference, however, is the scale that this occurs on: YouTube is undeniably the larger service. This is true from every angle we have investigated, not least its infrastructure, which contains many thousands of servers [Calder et al. 2013] compared to just 173 observed within our traces of YouPorn (§ 4). More noteworthy is the difference in upload rates though. In total, YouTube receives 300 hours of video uploaded every minute [YouTube 2015], dwarfing YouPorn’s 78 uploads per day (§ 5.2). Although this has increased to 140 per day during the last three months of the trace, it is still tiny compared to YouTube. These numbers are most interesting when contrasted with the number of page visits each website receives. YouTube reported in 2008 to have well over 140 million videos, whilst YouPorn’s corpus had just 183,639 ($\frac{1}{760}$ of YouTube’s corpus size). Despite this, YouTube gets just 100 times more page views than YouPorn does [Alexa 2014]. In other words, YouPorn has a much more concentrated and popular corpus: 93% of the videos receive at least 10k views. This can be compared to Vimeo, where only 1.9% of videos generate over 10k views [Sastry 2012]. This denser distribution of views is also represented in the popularity skew: the top 10% of YouPorn videos garner only 65% of the total views, whilst this value is 80% for YouTube [Cha et al. 2009] and 82% for Vimeo [Sastry 2012]. In other words,

Table III. Comparison of YouTube vs. YouPorn.

Attribute	YouTube	YouPorn
Content Type	Generic content, including professional and amateur uploads	Adult Content, primarily professional uploads (36% uploaded by YouPorn itself, with top 100 uploaders all professional)
Content Duration	Mean of 264 seconds [comSCORE 2014]	Mean of 526 seconds; median of 354 seconds
Upload Rate	300 hours of video uploaded every minute [YouTube 2015]; 443,653 over 5 years in Vimeo [Sastry 2012]	Mean of 78 videos per day
Number of Uploaders	Unknown, but over 1 billion registered users [YouTube 2015]	5849 registered uploaders, and 33k anonymous uploads
Number of visits	4 billion visits per day in 2012 [Kupka 2010]	4.85 billion visits in 2012, mean of 13.3 million visits per day [YouPorn 2012a]
Popularity Skew	Top 10% of videos garner 80% of views for YouTube [Cha et al. 2009] and 82% for Vimeo [Sastry 2012]	Top 10% of videos garner 65% of views
Views per Video	Mean 11k [Chatzopoulou et al. 2010]; mean Entertainment category 2197, Science and Technology category 2140 [Cha et al. 2009]; Mean of 1467 per video in Vimeo [Sastry 2012]	Mean of 335k; median of 84.7k
Number of Comments	Mean of 475 comments per video [Siersdorfer et al. 2010]	Mean of 5.55 comment per video; median of 1
% Comments per View	0.16% [Cha et al. 2009]; 0.17% [Chatzopoulou et al. 2010]	0.0017%
Number of Ratings	Mean of 21.95 ratings per video [Chatzopoulou et al. 2010]; mean of 7.2 in Vimeo [Sastry 2012]	Mean of 515.38; median of 120
% Ratings per View	0.22% [Cha et al. 2009] 0.199% [Chatzopoulou et al. 2010]	0.15%
Average Rating	Mean of 3.51/5 [Chatzopoulou et al. 2010]	Mean of 74.17/100 (≈3.71/5); median of 77
Number of Categories	15 categories	62 categories
Category Co-location	Limit of 1 category per video	Mean 3.47 per video; median of 3

views are more evenly spread across adult content (§ 6). This is driven by several reasons, however, most notable is the greater flexibility that people have when selecting what to watch on YouPorn (§ 9). This, when combined with the smaller corpus, enables users to spread their views far more evenly across videos.

Other issues that affect the above popularity trends include a lack of external referrals, and a collection of short clips that are easy to browse (§ 9). Both YouTube and YouPorn have a bias towards these short videos. 97.9% of YouTube videos were found to be below 10 minutes [Cheng et al. 2008]. This is largely due to the 10 minute duration limit (since lifted to 15 minutes) that YouTube places on normal accounts.[10] YouPorn does not implement such limits, yet we find that 72% are below 10 minutes. In 2014, Comscore reported that the average YouTube video was 264 seconds [comSCORE 2014], compared to 526 for YouPorn. Hence, YouPorn tends to have longer videos (§ 5.1). Despite this, both services have a propensity for short videos, allowing users to "pick and mix" their content quite conveniently.

A key driving force behind the differences observed is the nature of individual uploaders in the two services. YouPorn is driven by a much smaller set of active uploaders than we see in traditional UGC. Rather than the amateurs prevalent on YouTube, we find that the bulk of uploaders are commercial entities in YouPorn. Although YouTube has more recently tried to engage with commercial providers, YouPorn has long relied on such agreements. In fact, we found that 36% of all content had been uploaded by YouPorn itself. This is likely driven by the distinct business models: whereas

[10]Users can apply for an account that permits longer uploads.

YouTube monetises through advertising, YouPorn relies more heavily on content partnerships and revenue sharing with professional studios [YouPorn 2015]. This raises another key difference, relating to their operating environments. Adult content services operate in a highly fragmented marketplace, whilst YouTube has gained clear dominance, far beyond its smaller competitors, e.g., Vimeo. As such, adult video provision is spread across a number of high profile websites with billions of views accumulated by other services such as PornHub [Tyson et al. 2015].

The third thing we investigated was the "2.0" features of YouPorn (§ 7). Again, we found distinct differences in their usage. Whereas for YouTube, there is a comment for every 204 views [Thelwall et al. 2012], in YouPorn it is only one for every ≈39k views overall (§ 7.1). From a study of 67,290 videos discovered via searching for popular keywords, Siersdorfer *et al.* found an average of 475 comments per video [Siersdorfer et al. 2010], whilst Chatzopoulou *et al.* found a much lower value of 18.98 comments per video [Chatzopoulou et al. 2010]. Either way, comments are considerably less frequent on YouPorn, with a mean of just 5.5 comments per video, and 37% receiving no comments whatsoever.[11] Different trends, however, are seen with ratings. YouTube videos have, on average, 21.95 ratings per video [Chatzopoulou et al. 2010]. This can be contrasted with YouPorn, which has a much higher mean of 515.38 ratings per video, with approximately 90% of videos being rated more than 25 times (§ 7.2). This is largely driven by YouPorn's smaller, more densely popular, corpus. As such, proportionally, YouTube still converts a greater number of views into comments and ratings. In terms of averages, YouTube videos get 11k views, of which it converts 18.98 (0.17%) into comments and 21.95 (0.199%) into ratings [Chatzopoulou et al. 2010]. Similar figures are reported elsewhere: 0.16% and and 0.22% [Cha et al. 2009]. This is quite significantly higher than YouPorn, which only manages equivalent conversion rates of 0.0017% and 0.15% respectively. By any measure, YouTube viewers are more engaged in web 2.0 activities. This could be due to any of a number of reasons, not least the nature of YouPorn lending itself less to online interaction around content. For example, generic platforms like YouTube receive a wide range of content types, including those on polarising topics such as politics and religion, which incite more user interaction [Siersdorfer et al. 2010]. That said, it is worth noting that the actual ratings achieved are relatively similar; YouTube videos have an average rating of 3.51/5 [Chatzopoulou et al. 2010] compared to 74.17/100 (≈3.71/5) for YouPorn.

Of course, YouTube is not the only online video repository; we use it here as it is the most dominant player in the user-generated market. However, there are many others UGC platforms (e.g., DailyMotion, GorillaVid), as well as portals focussing on professional content (e.g., Netflix, Hulu). Direct comparisons with these are beyond the scope of this paper, although we wish to briefly note that, in terms of scale, Porn 2.0 portals are amongst the most popular in Alexa.

11. CONCLUSION

This paper has presented a detailed measurement study of YouPorn. Five key aspects of this system have been inspected: the delivery infrastructure, the video upload characteristics, the nature and evolution of content popularity, the use of 2.0 features and the impact of using categories. Our measurements show that YouPorn is a popular provider, with billions of views being accumulated for its videos. Despite being heavily promoted as a UGC website, we also discover a notable commercial element to its content, as well as a seemingly well managed vetting procedure. Several other interesting characteristics were also observed, particularly relating to the fast decay of content popularity as measured from the perspective of views and ratings, as well as the dependency users have on category metadata. Investigation uncovered the main reason behind these observations: the predominant use of YouPorn's browsing options. This is driven by the flexibility that many users have when accessing adult media: they do not seek a specific video, rather, they search for *any* video that falls with certain interest constraints. We therefore propose to exploit this observation by shaping users' browsing behaviour towards videos with a low network cost. Our future work involves exploring

[11] Note that the sampling of [Siersdorfer et al. 2010] and [Chatzopoulou et al. 2010] could introduce bias that our study does not suffer from.

this possibility further, as well as deep diving into other network aspects (e.g., traffic quantities). We are particularly keen to acquire more detailed session-level traces to better understand how individual preferences emerge. Ideally, this would include longitudinal data, providing insight into how user behaviour evolves over time. Such data could be used to inform superior caching and delivery decisions, as well as offering information on fine-grained interactions such as video playback skipping.

REFERENCES

Henrik Abrahamsson and Mattias Nordmark. 2012. Program popularity and viewer behaviour in a large TV-on-demand system. In *Proc. IMC*. ACM, New York, NY, USA, 199–210. DOI:http://dx.doi.org/10.1145/2398776.2398798

Alexa. 2014. Alexa Website Rankings. http://www.alexa.com/. (2014).

Sebastian Anthony. 2012. Just how big are porn sites? http://www.extremetech.com/computing/123929-just-how-big-are-porn-sites. (April 2012).

John G Apostolopoulos, Wai-tian Tan, and Susie J Wee. 2002. Video streaming: Concepts, algorithms, and systems. *HP Laboratories, report HPL-2002-260* (2002).

Feona Attwood. 2010. *Porn.com: Making sense of online pornography*. Vol. 48. Peter Lang.

Youmna Borghol, Sebastien Ardon, Niklas Carlsson, Derek Eager, and Anirban Mahanti. 2012. The untold story of the clones: content-agnostic factors that impact YouTube video popularity. In *Proc. of ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1186–1194.

Andrew Brampton, Andrew MacQuire, Michael Fry, Idris A Rai, Nicholas JP Race, and Laurent Mathy. 2009. Characterising and exploiting workloads of highly interactive video-on-demand. *Multimedia systems* 15, 1 (2009), 3–17.

Matt Calder, Xun Fan, Zi Hu, Ethan Katz-Bassett, John Heidemann, and Ramesh Govindan. 2013. Mapping the expansion of Google's serving infrastructure. In *Proceedings of the 2013 conference on Internet measurement conference*. ACM, 313–326.

Jason S Carroll, Laura M Padilla-Walker, Larry J Nelson, Chad D Olson, Carolyn McNamara Barry, and Stephanie D Madsen. 2008. Generation XXX Pornography Acceptance and Use Among Emerging Adults. *Journal of adolescent research* 23, 1 (2008), 6–30.

Meeyoung Cha, Haewoon Kwak, Pablo Rodriguez, Yong-Yeol Ahn, and Sue Moon. 2009. Analyzing the Video Popularity Characteristics of Large-scale User Generated Content Systems. *IEEE/ACM Trans. Netw.* 17, 5 (2009), 1357–1370.

Meeyoung Cha, Pablo Rodriguez, Jon Crowcroft, Sue Moon, and Xavier Amatriain. 2008. Watching television over an IP network. In *Proc. IMC*. ACM, 71–84.

Gloria Chatzopoulou, Cheng Sheng, and Michalis Faloutsos. 2010. A First Step Towards Understanding Popularity in YouTube. In *INFOCOM Workshops*. IEEE, 1–6. DOI:http://dx.doi.org/10.1109/INFCOMW.2010.5466701

Xu Cheng, Cameron Dale, and Jiangchuan Liu. 2008. Statistics and social network of YouTube videos. In *16th International Workshop on Quality of Service (IWQoS)*. IEEE, 229–238.

Cisco. 2015. Cisco Visual Networking Index: Forecast and Methodology, 2014-2019. (May 2015).

comSCORE. 2014. Online Video Rankings. http://www.comscore.com/Insights/Press-Releases/2014/2/comScore-Releases-January-2014-US-Online-Video-Rankings. (2014).

Al Cooper. 1998. Sexuality and the Internet: Surfing into the new millennium. *CyberPsychology & Behavior* 1, 2 (January 1998), 187–193.

Jonathan Coopersmith. 2006. Does Your Mother Know What You Really Do? The Changing Nature and Image of Computer-Based Pornography. *History and Technology* 22, 1 (2006), 1–25.

Riley Crane and Didier Sornette. 2008. Robust dynamic classes revealed by measuring the response function of a social system. *Proc. of the National Academy of Sciences* 105, 41 (2008), 15649–15653.

Sally Jo Cunningham and David M Nichols. 2008. How people find videos. In *Proc. of the 8th ACM/IEEE-CS joint conference on Digital libraries*. ACM, 201–210.

Kristian Daneback, Anna Sevcikova, Sven-Axel Månsson, and Michael W Ross. 2012. Outcomes of using the internet for sexual purposes: fulfilment of sexual desires. *Sexual Health* 10 (December 2012), 26–31. Issue 1. DOI:http://dx.doi.org/10.1071/SH11023

Gail Dines. 2010. *Pornland: How Porn Has Hijacked Our Sexuality*. Beacon Press.

Yehia Elkhatib, Rebecca Killick, Mu Mu, and Nicholas Race. 2014. Just browsing? Understanding user journeys in online TV. In *Proc. of the ACM Conference on Multimedia*. 965–968. DOI:http://dx.doi.org/10.1145/2647868.2654980

Lei Guo, Enhua Tan, Songqing Chen, Zhen Xiao, and Xiaodong Zhang. 2008. The stretched exponential distribution of internet media access patterns. In *Proc. PODC*. ACM, 283–294.

Xiaojun Hei, Chao Liang, Jian Liang, Yong Liu, and Keith W Ross. 2007. A measurement study of a large-scale P2P IPTV system. *IEEE Trans. Multimedia* 9, 8 (2007), 1672–1687.

Weiming Hu, Ou Wu, Zhouyao Chen, Zhouyu Fu, and S. Maybank. 2007. Recognition of Pornographic Web Pages by Classifying Texts and Images. *IEEE Trans. Pattern Analysis and Machine Intelligence* 29, 6 (June 2007), 1019–1034. DOI:http://dx.doi.org/10.1109/TPAMI.2007.1133

Ryan Hurley, Swagatika Prusty, Hamed Soroush, Robert J. Walls, Jeannie Albrecht, Emmanuel Cecchet, Brian Neil Levine, Marc Liberatore, Brian Lynn, and Janis Wolak. 2013. Measurement and Analysis of Child Pornography Trafficking on P2P Networks. In *Proc. World Wide Web Conference*.

Anna Kupka. 2010. YouTube Reaches 4 Billion Video Views Per Day. http://www.forbes.com/sites/annakupka/2012/01/24/youtube-reaches-4-billion-video-views-per-day/. (Jan 2010).

Patricia G Lange. 2007. Commenting on comments: Investigating responses to antagonism on YouTube. In *Annual Conference of the Society for Applied Anthropology. Retrieved August*, Vol. 29. 2007.

Lassi A Liikkanen and Antti Salovaara. 2015. Music on YouTube: user engagement with traditional, user-appropriated and derivative videos. *Computers in Human Behavior* 50 (2015), 108–124.

Michael D. Mehta and Dwaine Plaza. 1997. Content Analysis of Pornographic Images Available on the Internet. *The Information Society* 13, 2 (1997), 153–161. DOI:http://dx.doi.org/10.1080/019722497129179

Gianfranco Nencioni, Nishanth Sastry, Gareth Tyson, Vijay Badrinarayanan, Dmytro Karamshuk, Jigna Chandaria, and Jon Crowcroft. 2015. SCORE: Exploiting global broadcasts to create offline personal channels for on-demand access. *IEEE/ACM Transactions on Networking* (2015).

Ogi Ogas and Sai Gaddam. 2011. *A billion wicked thoughts: what the world's largest experiment reveals about human desire*. Dutton.

Henrique Pinto, Jussara M Almeida, and Marcos A Gonçalves. 2013. Using early view patterns to predict the popularity of YouTube videos. In *Proc. of international conference on Web search and data mining*. ACM, 365–374.

Nishanth Sastry. 2012. How to tell head from tail in User-generated content corpora. In *Proc. Conference on Weblogs and Social Media*. Dublin, Ireland.

Michael Schuhmacher, Cäcilia Zirn, and Johanna Völker. 2013. Exploring YouPorn Categories, Tags, and Nicknames for Pleasant Recommendations. In *Proc. Workshop on Search and Exploration of X-Rated Information (SEXI 2013)*. 27–28.

H. Schulze and K. Mochalski. 2009. Internet Study 2007-2009. http://www.ipoque.com/resources/internet-studies/. (2009).

Stefan Siersdorfer, Sergiu Chelaru, Wolfgang Nejdl, and Jose San Pedro. 2010. How useful are your comments?: analyzing and predicting YouTube comments and comment ratings. In *Proceedings of the 19th international conference on World wide web*. ACM, 891–900.

John Suler. 2004. The online disinhibition effect. *Cyberpsychology & Behavior* 7, 3 (2004), 321–326.

Gabor Szabo and Bernardo A. Huberman. 2010. Predicting the popularity of online content. *Commun. of ACM* 53, 8 (Aug 2010), 80–88. DOI:http://dx.doi.org/10.1145/1787234.1787254

Mike Thelwall, Pardeep Sud, and Farida Vis. 2012. Commenting on YouTube videos: From Guatemalan rock to el big bang. *Journal of the American Society for Information Science and Technology* 63, 3 (2012), 616–629.

Gareth Tyson, Yehia Elkhatib, Nishanth Sastry, and Steve Uhlig. 2013. Demystifying Porn 2.0: A Look into a Major Adult Video Streaming Website. In *Proceedings of the 2013 Conference on Internet Measurement Conference*.

Gareth Tyson, Yehia Elkhatib, Nishanth Sastry, and Steve Uhlig. 2015. Are People Really Social on Porn 2.0?. In *Ninth International AAAI Conference on Web and Social Media*.

Mirjam Wattenhofer, Yannet Interian, Jon Vaver, and Tom Broxton. 2010. Catching a Viral Video. In *Proc. Workshop on Social Interaction Analysis and Service Providers (SIASP), colocated with ICDM*.

Gilbert Wondracek, Thorsten Holz, Christian Platzer, Engin Kirda, and Christopher Kruegel. 2010. Is the Internet for porn? An insight into the online adult industry. In *Proc. Workshop on Economics of Information Security*.

YouPorn. 2012a. YouPorn 2012: BIG Numbers, HARD Facts. http://blog.youporn.com/youporn-2012-big-numbers-hard-facts/. (2012).

YouPorn. 2012b. Youporn.com is now a 100% Redis Site. https://groups.google.com/forum/?fromgroups=#!topic/redis-db/d4QcWV0p-YM. (2012).

YouPorn. 2014. YouPorn. http://www.youporn.com. (2014).

YouPorn. 2015. YouPorn Content Partnership. http://www.youporn.com/contentpartnerprogram/makemoney/. (2015).

YouTube. 2015. YouTube Statistics. http://www.youtube.com/yt/press/en-GB/statistics.html. (2015).

Hongliang Yu, Dongdong Zheng, Ben Y Zhao, and Weimin Zheng. 2006. Understanding user behavior in large-scale video-on-demand systems. In *ACM SIGOPS Operating Systems Review*, Vol. 40. ACM, 333–344.

Michael Zink, Kyoungwon Suh, Yu Gu, and Jim Kurose. 2009. Characteristics of YouTube network traffic at a campus network–measurements, models, and implications. *Computer Networks* 53, 4 (2009), 501–514.