Characterizing Video Access Patterns in Mainstream Media Portals

Lucas C. O. Miranda[1,2] Rodrygo L. T. Santos[1] Alberto H. F. Laender[1]

[1]Departamento de Ciência da Computação
Universidade Federal de Minas Gerais
Belo Horizonte, MG, Brazil

[2]Samba Tech
Belo Horizonte, MG, Brazil

{lucmir,rodrygo,laender}@dcc.ufmg.br

ABSTRACT

Watching online videos is part of the daily routine of a considerable fraction of Internet users nowadays. Understanding the patterns of access to these videos is paramount for improving the capacity planning for video providers, the conversion rate for advertisers, and the relevance of the whole online video watching experience for end users. While much research has been conducted to analyze video access patterns in user-generated content (UGC), little is known of how such patterns manifest in mainstream media (MSM) portals. In this paper, we perform the first large-scale analysis of video access patterns in MSM portals. As a case study, we analyze interaction logs across a total of 38 Brazilian MSM portals, including six of the largest portals in the country, over a period of eight weeks. Our analysis reveals interesting static and temporal video access patterns in MSM portals, which we compare and contrast to the access patterns reported for UGC websites. Overall, our analysis provides several insights for an improved understanding of video access on the Internet beyond UGC websites.

Categories and Subject Descriptors

H.3.5 [**Online Information Services**]: Web-based services

General Terms

Human Factors, Measurement

Keywords

Online Video, Video Access Patterns, User-Generated Content, Mainstream Media, Temporal Analysis

1. INTRODUCTION

Online video is changing the way people interact and collaborate on the Web. Today, users can easily share a high-quality video directly from any Internet-enabled device with millions of viewers. On the other hand, Internet video consumption has risen drastically over the last years. Indeed, a report from June 2010 estimated that 69% of all Internet users watch or download videos online, while 14% have posted videos [12]. Taking advantage of this trend, people have been adopting it in many applications, like product and service promotion, advertising, marketing campaigns, online coursers, among others. For businesses, this presents an opportunity to make information sharing more attractive and to reach out to the right audience.

Online video can be classified according to its providers. In particular, mainstream media (MSM) comprises videos shared by specialized providers (like online press, news portals, and entertainment portals) in their own website. In turn, user-generated content (UGC) includes videos shared by Internet users usually through video sharing websites or social media, such as YouTube[1] and Vimeo.[2] Much research has been devoted to analyzing the patterns of interaction between users and the videos posted in UGC websites (e.g., [2, 4, 5, 14]). In particular, UGC videos are known to be typically short and biased towards categories such as "Music" and "Comedy" [5]. On the other hand, little is known about video access patterns in MSM portals, primarily since detailed interaction logs from such portals are not generally available or are restricted to a few players in the online video sharing market, such as content distributors.

In this paper, we present an analysis of access patterns in Brazilian MSM portals, collected in association with the largest online video distribution platform in Latin America. Brazil is one of the 10 largest online video markets worldwide. Brazilian online viewing audience reached 43 million unique viewers in December 2012.[3] In addition, the online video penetration in Brazil peaked at 82% of its Internet users. Despite the increasing figures, there is still margin for growth as video penetration is slightly below the worldwide average of nearly 84%.

Our analysis is based on aggregate data from a total of 38 Brazilian MSM portals, including six of the largest portals in the country, and reveals interesting static and temporal access patterns. Regarding static patterns, in line with previous research, we observe that the distribution of views per video presents a truncated long tail, indicating a lower-than-expected amount of unpopular videos. In contrast, MSM portals typically have much longer videos than UGC websites. In addition, the most popular video categories in UGC websites—"Music" and "Comedy"—are not as dominant in MSM portals. Regarding temporal patterns, we observed cyclic hourly and daily access patterns to MSM videos over our considered 8-week time frame. We also noted the preva-

WWW 2013 Companion, May 13–17, 2013, Rio de Janeiro, Brazil.
ACM 978-1-4503-2038-2/13/05.

[1]http://www.youtube.com
[2]http://www.vimeo.com
[3]http://bit.ly/WPpT4G

lence of very low values for video retention. Finally, studying how views evolve over time, we concluded that the average video life span depends on the video category but, in general, it is very short.

Thus, the main contributions of this paper are:

1. A thorough characterization of static and temporal video access patterns in major MSM portals;

2. A comparison of the observed patterns with those previously reported for UGC video-sharing websites.

The remainder of this paper is organized as follows. In Section 2, we overview the related literature on the analysis of online video sharing. In Section 3, we discuss the methodology underlying the analysis conducted in Section 4. Finally, in Section 5, we present our concluding remarks and provide some directions for future work.

2. RELATED WORK

The work presented in this paper relates to a broad spectrum of research on online video sharing, ranging from network traffic characterization for capacity planning at a lower level (e.g., [8, 9, 13, 15]) to social network analysis at a higher level (e.g., [2, 3, 6, 7, 11, 14]). Specifically, our work can be better placed somewhere in between these two extremes, by focusing on the characterization of video access patterns at the application (as opposed to the network) level.

Acharya et al. [1] were among the first to analyze video access patterns on the Web, in a period when online video watching was not as widespread as it is today. In particular, they analyzed the access patterns to 139 videos from a small-scale experiment in a Swedish university over a 6-month period between 1997 and 1998. As a result of this analysis, they observed cyclic access patterns, with the video watching activity dropping during weekends compared to weekdays. They also observed a temporal locality in the access patterns, and a low retention for 45% of all videos, the majority of which being stopped after only 5% of their duration. In Sections 3 and 4, we conduct a similar temporal analysis, however at a much larger scale, covering millions of videos published by major Brazilian MSM portals.

In recent years, with the emergence of social media on the Web, several studies have been conducted to analyze video access patterns in UGC websites, notably YouTube (e.g., [4, 5]). For instance, Cheng et al. [5] analyzed a 3-month crawl of YouTube from early 2007, comprising a total of 2.6 million videos, 22.9% of which falling in the "Music" category, the most popular of YouTube's 12 defined categories. In their study, they observed that short clips make the majority of videos on YouTube, with 97.8% of all videos spanning less than 5 minutes. Regarding the distribution of views per video, their analysis showed that it follows a truncated power-law rather than a standard Zipf's distribution [10], indicating a lower-than-expected amount of unpopular videos (i.e., the long tail). In addition, by tracking a total of 43 thousand videos for a period of seven weeks, they observed that a total of 70% have a decaying popularity growth as time passes, denoting a short life-span.

Regarding the comparison of UGC and non-UGC content, Cha et al. [4] performed a large-scale analysis of content production on YouTube, in contrast to specialized non-UGC video providers, such as Netflix, LOVEFiLM, and Yahoo! Movies. Their analysis revealed a substantially larger growth rate of content production in UGC compared to non-UGC websites. On the other hand, the distribution of videos per publisher between the two scenarios showed a similar power-law behavior, with the exception that the median production of a non-UGC publisher (e.g., a movie director) is naturally capped by the temporal and monetary constraints imposed by the movie industry. Their analysis also showed that the median duration of a UGC video is two orders of magnitude shorter than that of a non-UGC video. In this paper, we also analyze video access patterns in non-UGC websites. However, to the best of our knowledge, ours is the first analysis focused on general-purpose MSM portals as opposed to specialized non-UGC video providers. In addition, we analyze both the content production and consumption on MSM portals, as will be discussed in Sections 3 and 4.

3. EXPERIMENTAL METHODOLOGY

To understand the behavior of users accessing MSM portals, as a case study, we have analyzed records of Internet users interacting with some of the leading Brazilian MSM portals. The data for our case study was collected in association with Samba Tech.[4] Samba Tech is the largest online video platform in Latin America, providing solutions in video management and distribution for customers including large media groups, soccer teams, news websites, educational institutes and e-commerce channels in Brazil.

Samba Tech provides to its customers a platform for online video hosting, management, and distribution. Websites using the platform can also adopt Samba Tech's embedded player. We have collected user interactions while accessing videos delivered by Samba Tech during a period of 8 weeks, from June to August, 2012. The data was collected by a script in Samba Tech's embedded player. Every user action on the player in a customer's web page results in an API call to a server. The server aggregates all events, from different websites, in a database.

For each user accessing a video at a given time, we have logged what we call a *session*. In particular, a session comprises all user interactions with the player (e.g., play, pause, resume, and video progress) and general information about the video being watched (e.g., category, publication date, and duration). We also converted all access times, originally logged as GMT, to the local time zone. All user data was collected entirely anonymously. The video content and the video provider's identity were also anonymized during the collection process.

Start date	Jun 24th, 2012 (Sun)
End date	Aug 18th, 2012 (Sat)
Unique sessions	110,626,789
Unique users	43,217,621
Unique videos	127,068
Video duration (mean)	433.5s
Video duration (s.d.)	782.9s

Table 1: Salient statistics of our analyzed dataset.

Table 1 describes the salient statistics of the collected dataset. The dataset includes user interactions recorded continuously over a period of 8 weeks, from June 24th, 2012 (Sun) to August 18th, 2012 (Sat), so as to enable the observation of both static as well as temporal factors over a

[4]http://www.sambatech.com

long timespan. Moreover, it is noteworthy that the collected dataset comprises the entire log of the user interactions recorded over the duration of this study, as opposed to a sample of it. In particular, the dataset comprises around 110 million sessions, 43 million unique users and 127 thousand unique videos, with a mean duration of around 7 minutes. The relatively high standard deviation of video durations reveals the heterogeneity of the content posted on the considered general-purpose MSM portals. As we will show in Section 4.1, such a high variability is primarily due to the presence of videos from multiple distinct categories.

4. EXPERIMENTAL ANALYSIS

In this section, we perform a thorough analysis of the interaction logs collected from major Brazilian MSM portals, as described in Section 3. In particular, this analysis aims to answer the following research questions:

Q1. Which access patterns emerge from analyzing a static snapshot of MSM portals aggregated over time?

Q2. Which temporal patterns can be inferred by analyzing user interactions at different points in time?

To address these research questions, we perform two broad sets of analyses. In particular, in Section 4.1, we perform a static analysis of our collected dataset, in which the whole dataset is analyzed in aggregate, regardless of temporal aspects. In Section 4.2, we complement our static picture of Brazilian MSM portals with a temporal analysis, aimed to investigate user access patterns at different points across the 8-week timespan of our collected dataset.

4.1 Static Analysis

In this section, we address research question Q1, by analyzing a static snapshot of our 8-week dataset, in order to uncover its salient access patterns. In particular, we analyze the distribution of video access requests in light of various user and video properties, such as category and duration. In addition, we contrast the uncovered patterns to those reported in the literature for users interacting with videos on UGC websites.

4.1.1 Video Categories

The topical nature of a video is a determinant factor to understand its access patterns. In particular, the videos hosted in the Samba Tech platform can be assigned to one of nine fixed categories. Such an assignment is optionally performed by the video provider when the video is uploaded. Figures 1 and 2 show, respectively, the distribution of videos and views per category in our dataset.

From Figure 1, we first observe that 53.6% of all videos (68,098 from a total of 127,068) have not been assigned to any category. In turn, the categories with the most videos are: "Politics" (17.1%), "Entertainment" (14.3%), "Sports" (11.1%), and "Science" (2.3%). These categories are also popular in UGC video sharing websites, such as YouTube. On the other hand, while "Music" and "Comedy" are among the most popular categories on YouTube (22.9% and 12.1%, respectively) [6], they appear with relatively fewer assignments in our dataset (less than 1%). Arguably, such a discrepancy is justified by the absence of music providers in our dataset and by the presumable preference of UGC users to share music and comedy content.

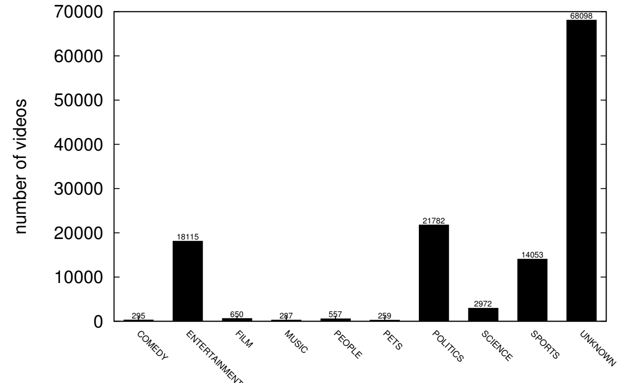

Figure 1: Number of videos per category.

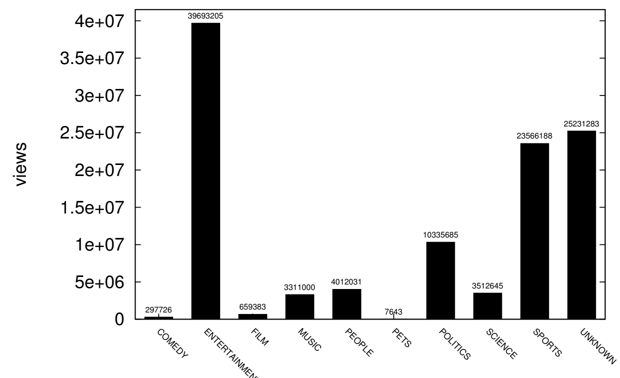

Figure 2: Number of views per category.

From Figure 2, we observe that the number of views per category does not necessarily correspond to the number of videos per category. Indeed, the category with the most videos, "Politics", received less views than categories such as "Entertainment" and "Sports". "Entertainment" was the category with the most views overall (35.9%), arguably because videos from this category usually have a strong "viral effect", tending to be more shared. Considering the ratio of views per video, the categories "Music" and "People" stand out from the others with 11,536.59 and 7,202.93, respectively.

4.1.2 Video Duration

Statistics reported for YouTube have shown that it is mostly comprised of short video clips (20.6% of the videos have less than one minute of duration and 17.1% have 3 to 4 minutes) [6]. In contrast, as previously shown in Table 1, the mean video duration in our entire dataset is 433.5s, i.e., just above 7 minutes. Moreover, the mean duration of a video is closely related to its category, as shown in Figure 3. In the figure, mean duration is expressed in seconds, with error bars denoting a 95% confidence interval for the mean.

From Figure 3, we observe that videos from the "Pets" category and those without an assigned category have the longest duration (over 600s) compared to videos in other categories. "Entertainment" and "Politics" have also a long mean duration, probably because most of the videos in these categories are short documentaries, reports, or parts of entire TV shows. In contrast, "Comedy", "Music", and "Sports" comprise much shorter videos on average, corresponding to short news reports, ads, or music clips. The predominance

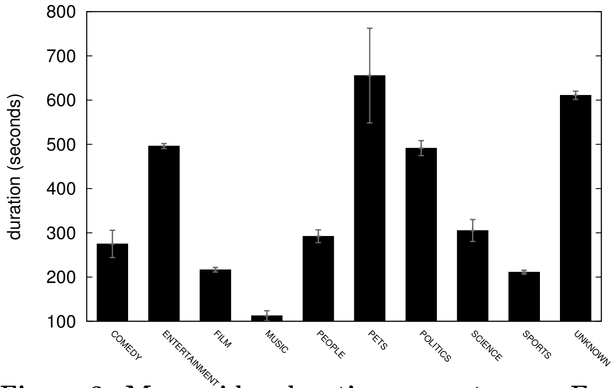

Figure 3: Mean video duration per category. Error bars denote 95% confidence intervals.

of short videos in these categories also resembles the scenario observed with YouTube, where these categories are among the most popular ones [6].

4.1.3 Views

There is no authentication requirement for most users accessing videos in the analyzed MSM portals. However, we can still keep track of which videos have been watched by a user over a period of time by resorting to cookies. Based upon this tracked data, Figure 4 shows the complementary cumulative distribution function (CCDF) of the number of views per user in our dataset.

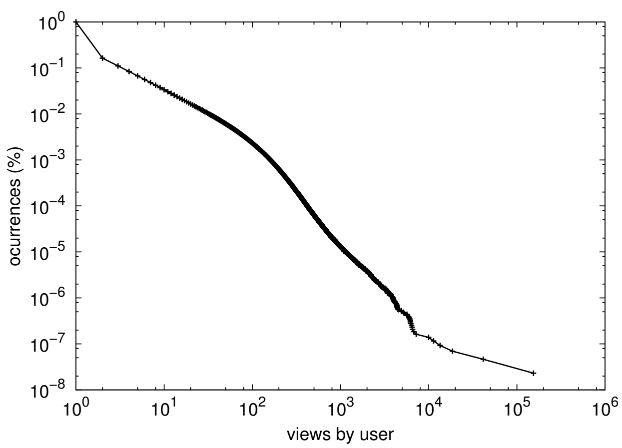

Figure 4: Number of views per user (CCDF).

From Figure 4, we observe that all users watched at least one video (i.e., the x and y axes start from 1). The CCDF distribution reveals a long tail of users that watched a small number of videos, whereas a small fraction of users watched substantially more videos. Indeed, less then 10% of the users watched at least ten videos and much less than 1% watched at least one hundred videos. Complementary to this observation, Figure 5 shows the CCDF of the number of views per video.

From Figure 5, we observe a less steep distribution, where more than 10% of all videos received at least one hundred views. In contrast, much fewer videos received more than a thousand views. While also showing a long tail of modestly watched videos, this distribution does not follow Zipf's law, in accordance with previous observations reported for YouTube [1, 6].

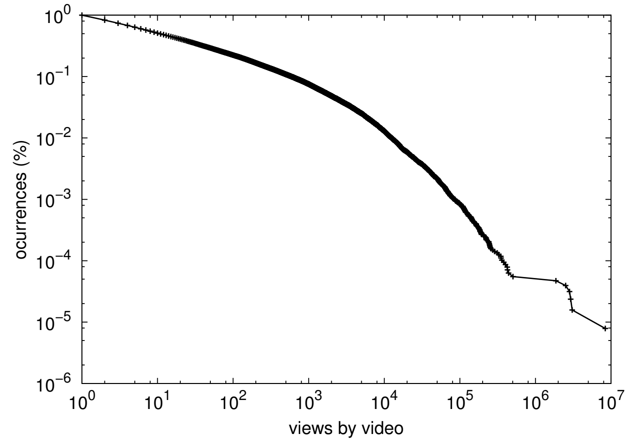

Figure 5: Number of views per video (CCDF).

In light of research question Q1, on the access patterns derived from a static snapshot of our 8-week dataset, the analysis in this section reveals an interesting parallel with previous studies on video access patterns in UGC websites, most notably YouTube. In particular, regarding the popularity of video categories, we observed a distinguishingly lower prevalence of "Music" and "Comedy" videos in the analyzed MSM portals, the two most popular categories in YouTube [6]. We have also observed that the average video length in these MSM portals depends heavily on the video category, whereas the majority of YouTube videos is relatively much shorter, regardless of their category. Finally, in line with previous research [1, 6], we have observed a truncated long tail in the distribution of views per video, which does not follow Zipf's law.

4.2 Temporal Analysis

In the previous section, we analyzed our entire dataset in aggregate, considering only static aspects. In this section, we seek to answer research question Q2, by focusing on the analysis of access patterns over time. In particular, our analysis comprises three broad dimensions: the evolution of access patterns, video retention, and video life span. Note that this analysis is only made possible because of the large time frame covered by our collected dataset.

4.2.1 Access Patterns

The frequency and time when users watch videos online may vary according to their particular daily routine. Indeed, it is possible to uncover some common temporal access patterns by analyzing our interaction data over time. For instance, Figure 6 shows the number of views (accesses) per day over the entire analyzed interval of 8 weeks (56 days). The first day in the interval (June 24th, 2012) is a Sunday, whereas the last day (August 18th, 2012) is a Saturday. The mean number of views per day in the period is 1,848,972.21, with extreme values of 1,094,646 and 2,416,967.

From Figure 6, we first observe that there is a clear cyclic pattern of accesses to MSM videos. In particular, for the analyzed interval, the number of views during weekdays is almost twice as much as the daily views during weekends. We

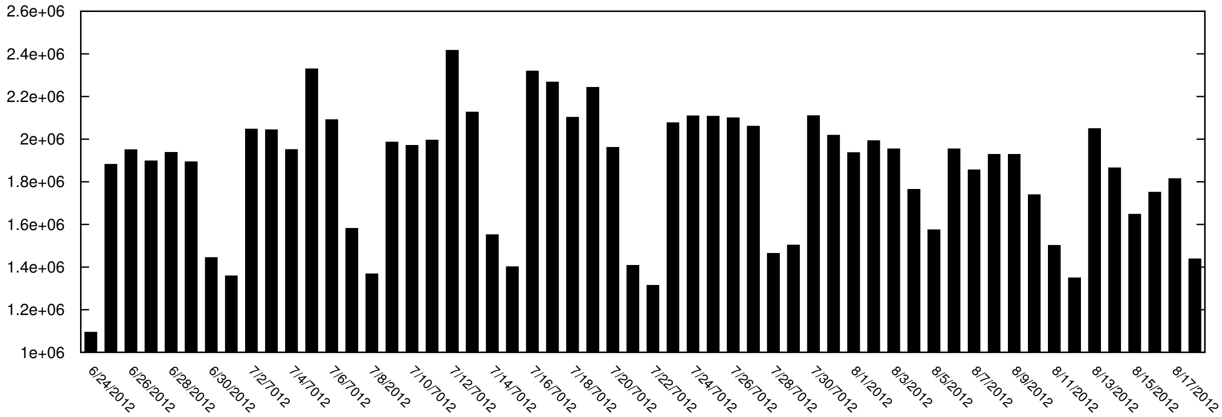

Figure 6: Number of views per day over 8 weeks.

can also observe that, with the exception of July 28th and 29th, more videos were viewed on Saturdays than on Sundays. Actually, Sunday is the day with the smallest average number of views during the week (1.370.790,87), whereas Thursday is the day with the largest average (2.087.578,37). To better understand access patterns over time, we have further analyzed the distribution of views at a finer granularity. To this end, Figure 7 shows the number of views per hour of the day. While we have analyzed the distribution for the entire dataset (8 weeks), for the sake of clarity, we show the results for a 7-day period between July 15th (Sunday) and 21st (Saturday), which represents a typical week in our dataset, with the other weeks showing a similar behavior.

From Figure 7, we observe that the distribution of views per hour of the day follows an expected and well-defined pattern. In particular, there is an accelerated rise in views from 7am to 12am. Then the number of views grows slowly, presenting some peaks and troughs. The highest peak usually occurs between 7pm and 8pm. After that, accesses decline quickly until about 6am, when they reach the lowest value. This pattern of access is observed invariably in other days.

Figure 8 shows the distribution of views per hour of the day for the four most popular categories. We can see that the common pattern shown in Figure 7 is also observed in the distributions of individual categories. However, it is possible to infer some specific patterns. For example, for category "Sports", the number of views is much higher on Monday and Thursday than on the other days of the week. This can be explained by the fact that the main soccer games in Brazil occur on Sunday and Wednesday, which result in more views on the immediately following days.

4.2.2 Retention

The number of accesses a video has received reflects its popularity, but not necessarily how much a user liked it. To estimate the video rating for a user, we calculate the video *retention* as the time a user spends watching the video divided by the video duration. We use the session time—i.e., the time difference between the last logged session event and the first one—as an approximation of viewing time. Figure 9 shows the CCDF of retention values for all sessions. From the figure, we observe that only about 25% of all sessions have a retention greater than 0.1. In other words, most users watch less than 10% of the content of each video.

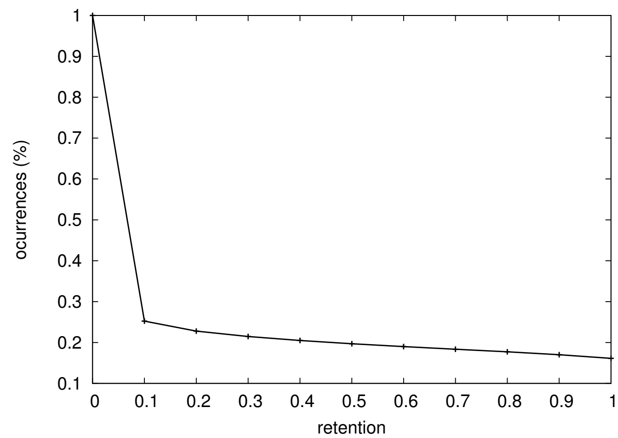

Figure 9: Video retention (CCDF).

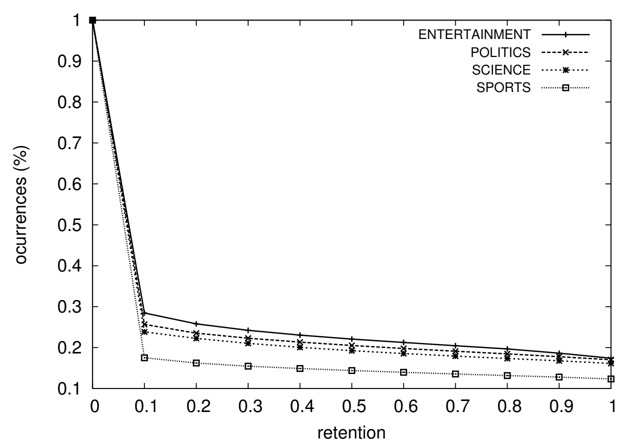

Figure 10: Video retention per category (CCDF).

We have also plotted the CCDF of video retention for the four most representative categories in Figure 10. From the figure, we note that all categories show similar retention distributions, differing slightly on the percentage of sessions with at least 10% of retention. Category "Sports" has less

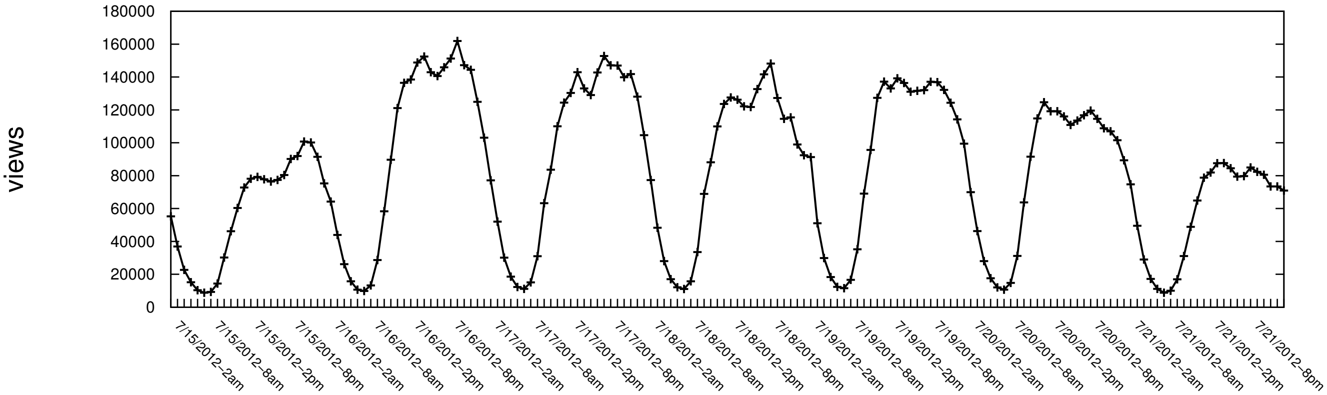

Figure 7: Number of views per hour of the day over 7 days.

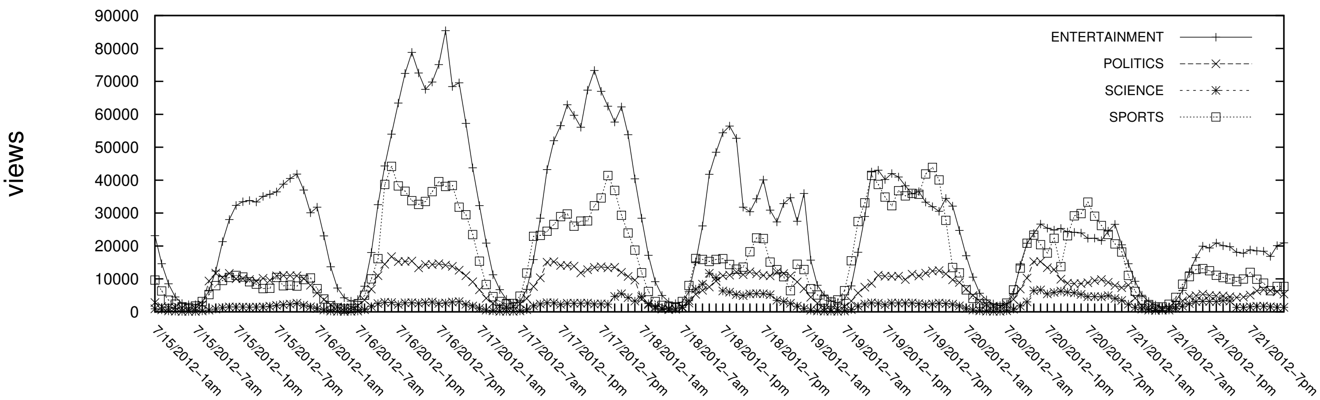

Figure 8: Number of views per hour of the day and category over 7 days.

than 20% of sessions with 0.1 of retention, while category "Entertainment" has almost 30% of sessions with at least 10% of retention.

4.2.3 Life Span

In Section 4.1.3, we analyzed the distribution of views aggregated over the whole interval of 8 weeks comprised by our dataset. In that analysis, only the total views received by a video (or the total views granted by a user) were considered. In this section, we examine how this amount of views is distributed over time.

To understand the evolution of views, it is necessary to track the accesses received by videos since their publication. To this end, we consider only videos that have been published in the first week of our collection (from June 24th to 30th), and track the number of views of these videos daily from the day of publication until the last day of our data collection (August 18th). Figure 11 shows the life span distribution (CDF) of these videos, measured in terms of the number of views received per day since the publication of each video.

From Figure 11, we observe that about 37% of the views received by a video occur within the same day of its publication (represented by 0 on the x axis), 67% occur until the end of the first day of publication, 85% are achieved before the 5th day, and 90% are achieved before the 9th day. Clearly, the evolution of views follows a logarithmic function. Most

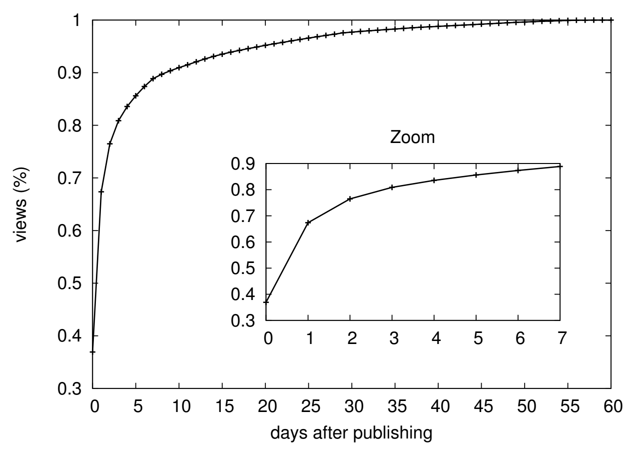

Figure 11: Video life span (CDF).

views occur as soon as a video is published, while less views are received on the subsequent days over a few weeks. In addition, only very few videos continue to be watched after one month of their publication, which indicates that the life span of MSM videos is usually very short.

A breakdown of the results in Figure 11 across the four most representative categories is shown in Figure 12, which displays the CDF of video life span per category. In addition, Figure 13 shows the same distribution for the first week only.

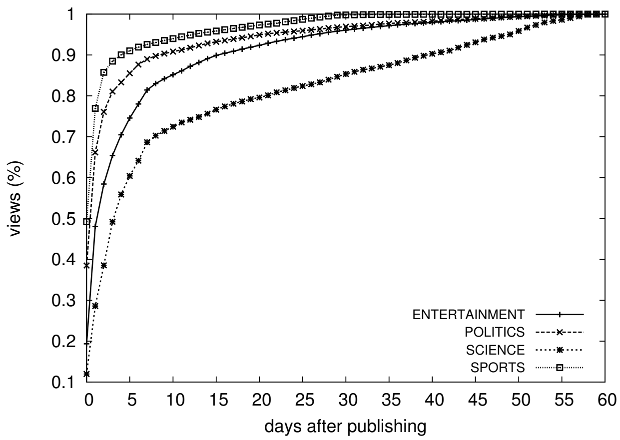

Figure 12: Video life span per category (CDF).

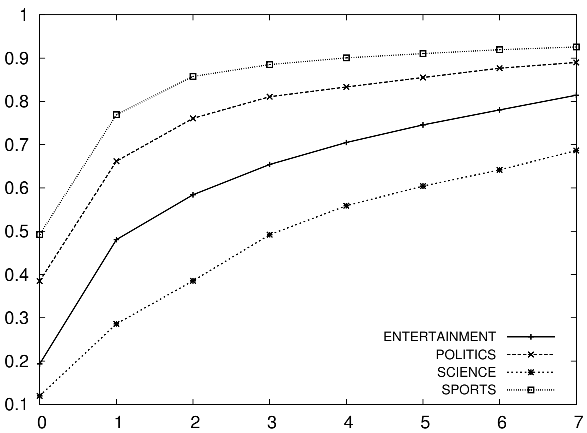

Figure 13: Video life span per category (CDF) over the first week.

From Figures 12 and 13, we observe the same logarithmic pattern exhibited in Figure 11. However, the most distinguishing difference is the convergence rate. In particular, about 77% of the views for videos from the category "Sports" occur on the first day of their publication, whereas videos from the category "Science" achieve only about 29% of the views in the same period. In fact, it takes almost 16 days for videos from the category "Science" to reach 77% of their total views.

In light of research question Q2, we have analyzed several video access patterns emerging over the time frame covered by our dataset. By analyzing views by day, we have observed that there is a cyclic access pattern and that the number of views during weekdays is much greater than during the weekends. A cyclic pattern also occurs in the daily distribution of views. We have also noted that the video retention varies slightly between categories but, in general, it takes low values for most users. Analyzing how views evolve over time, we conclude, as already reported by Cheng et al. [5] for YouTube videos, that the video life span is usually very short, with videos from some categories receiving most of their views on the first day after publication, while videos from other categories usually taking longer to reach the same percentage of their total accesses. Lastly, video providers can use this information to estimate how often content has to be released or to predict access statistics. It is also important to consider the differences between categories in how views evolve after a video publication.

5. CONCLUSIONS

In this paper, we have presented an extensive analysis of video access patterns in mainstream media (MSM) portals. Due to the limited availability of public data, little is known regarding access patterns to MSM videos. However, using data collected in association with Samba Tech, we could have a privileged perspective of user interactions while accessing some of the largest Brazilian MSM portals. In our analysis, we aimed to investigate which access patterns could be identified by considering a static snapshot of the collection and which temporal patterns could be inferred by observing user interactions over time.

Our static analysis revealed interesting patterns that can be compared with previous studies on video access patterns in UGC websites. We observed that the two most popular categories in YouTube are among the least prevalent in MSM portals. We have also noted that the average video duration in MSM portals depends heavily on the video category and that YouTube videos are relatively much shorter than MSM ones. Analyzing the distribution of views per video, we have observed a truncated long tail, which does not follow Zipf's law. Investigating the access to MSM videos over time, we have found some recurring patterns. In particular, a cyclic access pattern was observed by analyzing how views evolve by hour and by day. We have also noted the prevalence of very low values for video retention. Finally, studying how views evolve over time, we have concluded that video life span depends on video category but, in general, it is very short.

The access patterns revealed by static and temporal analyses have a broad applicability and can be used by MSM portals to improve service quality and enhance users' experience. The static investigation has brought general information about user behavior and preferences. The temporal analysis, on the other hand, may allow providers to better know when to publish and replace a given content. It is also possible to use this information to estimate users' reaction after content publishing. In fact, there are many applications that may benefit from our results, such as content-targeted advertising, recommendation, service personalization, etc. As future work, we plan to expand our analysis to investigate social aspects of MSM portals, such as how the users of such portals relate to one another, in terms of the similarity of their access patterns.

6. ACKNOWLEDGEMENTS

This work was partially funded by InWeb—The Brazilian National Institute of Science and Technology for the Web (grant MCT/CNPq 573871/2008-6) and by individual grants from CNPq. The authors would like to thank Samba Tech for granting them access to its video access logs.

7. REFERENCES

[1] S. Acharya, B. C. Smith, and P. Parnes. Characterizing user access to videos on the World Wide Web. In *Proceedings of the 7th SPIE Multimedia Computing and Networking Conference*, pages 130–141, 1999.

[2] F. Benevenuto, F. Duarte, T. Rodrigues, V. A. Almeida, J. M. Almeida, and K. W. Ross. Understanding Video Interactions in YouTube. In *Proceedings of the 16th ACM International Conference on Multimedia*, pages 761–764, 2008.

[3] F. Benevenuto, T. Rodrigues, V. A. F. Almeida, J. M. Almeida, and M. A. Gonçalves. Detecting spammers and content promoters in online video social networks. In *Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 620–627, 2009.

[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 *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, pages 1–14, 2007.

[5] X. Cheng, C. Dale, and J. Liu. Understanding the Characteristics of Internet Short Video Sharing: YouTube as Case Study. *CoRR*, abs/0707.3670, 2007.

[6] X. Cheng, C. Dale, and J. Liu. Statistics and Social Network of YouTube Videos. In *Proceedings of the 16th International Workshop on Quality of Service*, pages 229–238, 2008.

[7] X. Cheng and J. Liu. NetTube: Exploring Social Networks for Peer-to-Peer Short Video Sharing. In *Proceedings of the 28th IEEE International Conference on Computer Communications*, pages 1152–1160, 2009.

[8] P. Gill, M. Arlitt, Z. Li, and A. Mahanti. YouTube Traffic Characterization: A View from the Edge. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, pages 15–28, 2007.

[9] G. Gürsun, M. Crovella, and I. Matta. Describing and Forecasting Video Access Patterns. In *Proceedings of the 30th IEEE International Conference on Computer Communications*, pages 16–20, 2011.

[10] C. D. Manning and H. Schütze. *Foundations of Statistical Natural Language Processing*. MIT press, 1999.

[11] J. C. Paolillo. Structure and network in the YouTube core. In *Proceedings of the 41st Hawaii International Conference on System Sciences*, pages 156–156, 2008.

[12] K. Purcell. The state of online video. Technical report, Pew Internet & American Life Project, 2010.

[13] M. Saxena, U. Sharan, and S. Fahmy. Analyzing video services in Web 2.0: a global perspective. In *Proceedings of the 18th International Workshop on Network and Operating Systems Support for Digital Audio and Video*, pages 39–44. ACM, 2008.

[14] G. Szabo and B. A. Huberman. Predicting the popularity of online content. *Communications of the ACM*, 53(8):80–88, 2010.

[15] M. Zink, K. Suh, Y. Gu, and J. Kurose. Characteristics of YouTube network traffic at a campus network: measurements, models, and implications. *Computer Networks*, 53(4):501–514, 2009.