

Characterizing Video Access Patterns in Mainstream Media Portals

Lucas C. O. Miranda ^{1,2} Rodrygo L. T. Santos ¹ Alberto H. F. Laender ¹
{lucmir,rodrygo,laender}@dcc.ufmg.br

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

² Samba Tech
Belo Horizonte, MG, Brazil

Third Temporal Web Analytics Workshop (TempWeb03)
May 13, 2013



Outline

Introduction

Experimental Methodology

Experimental Analysis

Static Analysis

Temporal Analysis

Conclusions

Acknowledgements

Introduction

Online video is changing the way people interact and collaborate on the Web.

- ▶ 69% of all Internet users consume videos online.
- ▶ 14% have posted videos ¹.



¹The state of online video. Technical report, Pew Internet & American Life Project, 2010

Introduction

Online video consumption in Brazil:

- ▶ Among the Top 10 largest online video markets.
- ▶ 82% of users consume videos online.
- ▶ 43 million unique viewers.

	Total Unique Viewers	% Reach Web Population
Worldwide	1,279,264	83.8%
China	289,890	84.3%
US	188,130	84.9%
Japan	60,939	82.8%
Russia	55,591	90.6%
India	51,718	73.1%
Germany	47,617	82.9%
Brazil	42,998	82.2%
France	40,662	84.6%
UK	37,477	83.6%
Italy	23,857	83.0%

Top 10 Global Markets by Unique Video Viewers².

²<http://comscore.com>. February 6, 2013.

Introduction

Online video classification according to its providers:

- ▶ Mainstream Media (MSM).
- ▶ User-generated Content (UGC).



X



Introduction

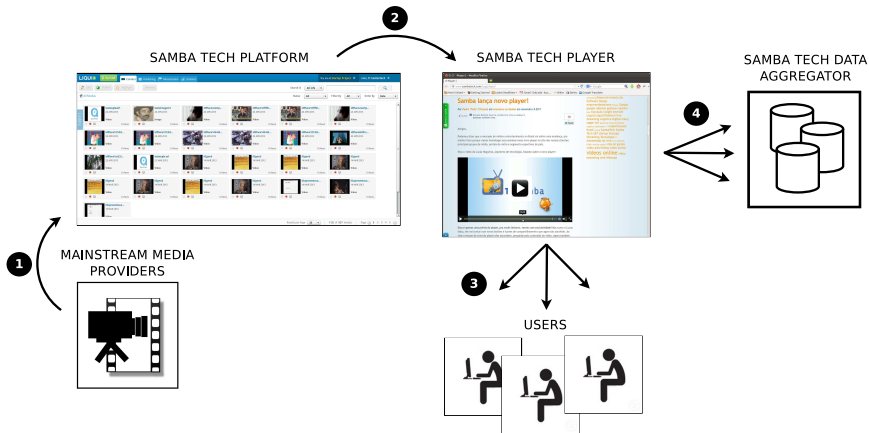
We present an analysis of access patterns in Brazilian MSM portals.

Motivation:

- ▶ The importance of online videos.
- ▶ The position of Brazil as one of the largest online video markets.
- ▶ Little is known about video access patterns in MSM portals, since detailed data is not generally available.

Experimental Methodology

Data was collected in association with Samba Tech ³



³ <http://sambatech.com>.

Experimental Methodology

Dataset information:

- ▶ Session(user, time) = player events + video info.
- ▶ 38 Websites. Including 6 of the largest Brazilian portals.

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

Salient statistics of our dataset.

Experimental Analysis

Research questions:

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

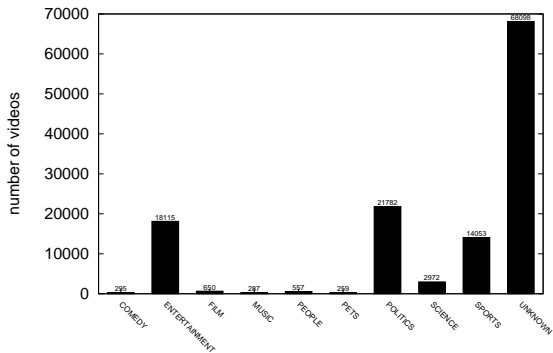
→ **Static Analysis**

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

→ **Temporal Analysis**

Static Analysis

Video Categories

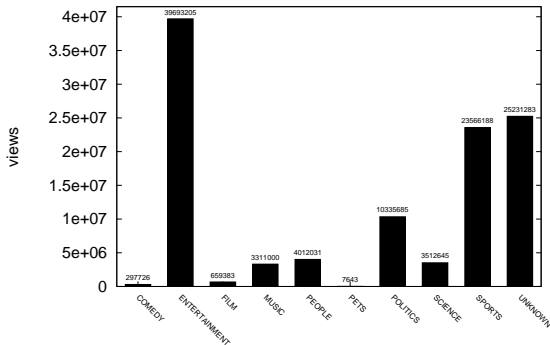


Number of videos per category.

- ▶ 53.6% have not been assigned to any category.
- ▶ Distribution similar to UGC Websites. But “Music” and “Comedy” are the two most present categories on YouTube.

Static Analysis

Video Categories

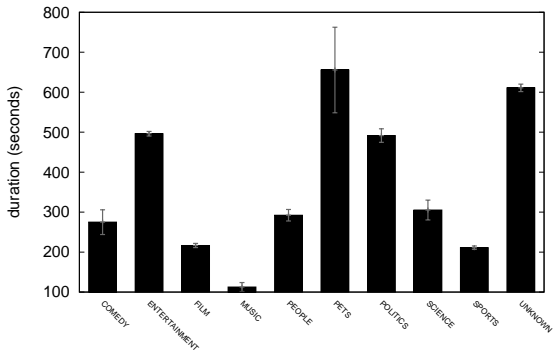


Number of views per category.

- ▶ There is not an exact correspondence with videos per category. “Politics” received less views than “Entertainment”.
- ▶ Considering views per video, “Music” and “People” stand out from the others.

Static Analysis

Video Duration



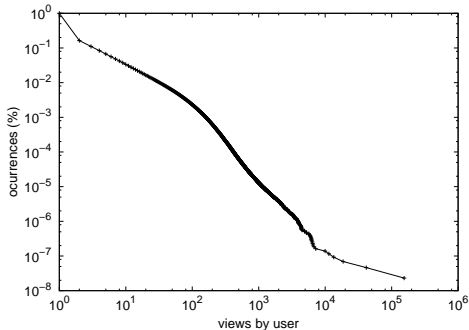
Mean video duration per category.

Error bars denote 95% confidence intervals.

- ▶ Video length depends heavily on the category.
- ▶ The majority of YouTube videos is relatively much shorter.

Static Analysis

Views

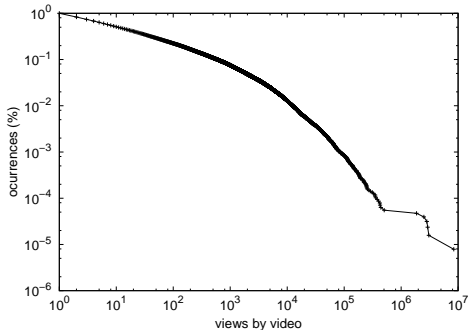


Number of views per user (CCDF).

- ▶ Long tail behaviour.
- ▶ Less than 10% of the users watched at least 10 videos.
- ▶ Much less than 1% watched at least 100 videos.

Static Analysis

Views

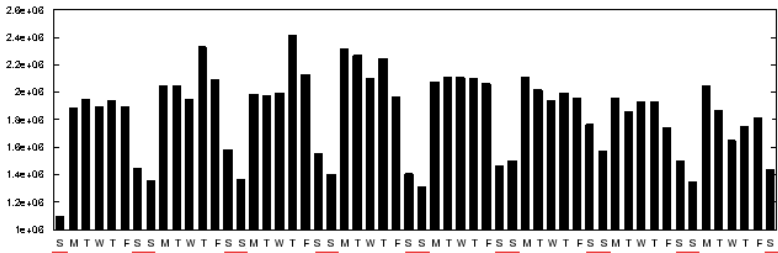


Number of views per video (CCDF).

- ▶ Long tail of modestly watched videos.
- ▶ But distribution does not follow Zipf's law (behaviour also reported for YouTube).

Temporal Analysis

Access Patterns

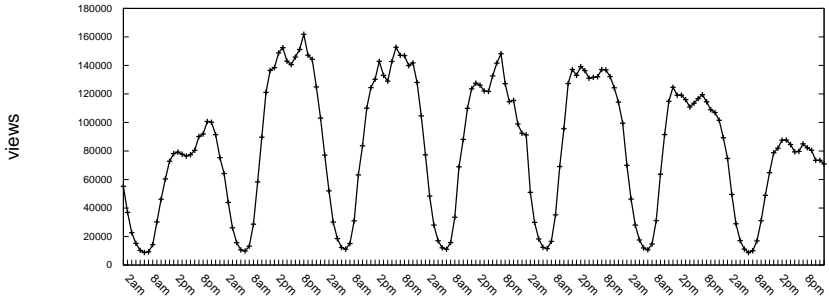


Number of views per day over 8 weeks.

- ▶ Cyclic pattern of accesses
- ▶ weekdays x weekends.

Temporal Analysis

Access Patterns

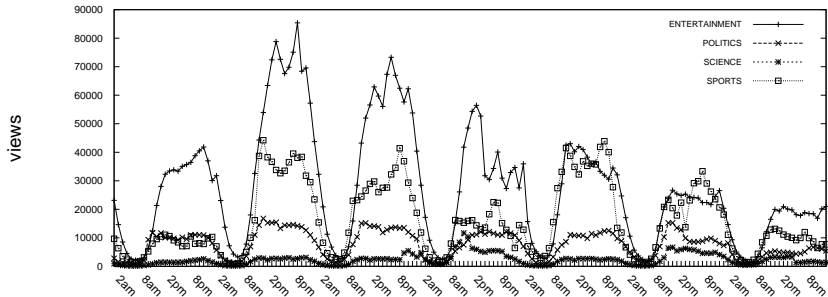


Number of views per hour of the day over 7 days.

- ▶ Expected and well-defined pattern.
- ▶ Highest peak usually occurs between 7pm and 8pm.
- ▶ Lowest value occurs about 6am.

Temporal Analysis

Access Patterns

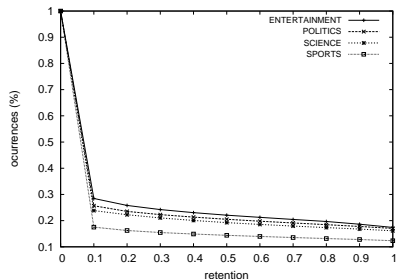
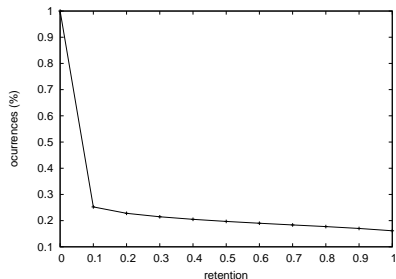


Number of views per hour of the day and category over 7 days.

- ▶ Follows the common behaviour, but there are specific patterns for each category.
- ▶ For example, for “Sports”, the number of views is much higher on Monday and Thursday.

Temporal Analysis

Retention

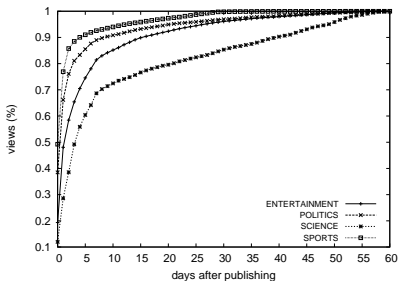
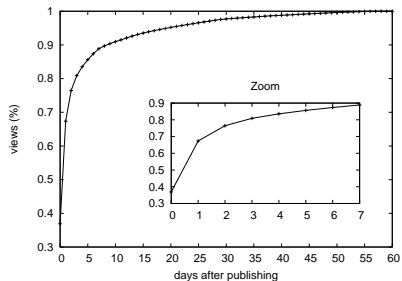


Video retention - total and per category (CCDF).

- ▶ Most users watch less than 10% of the video (low retention).
- ▶ Similar distributions between categories, differing slightly on the percentage of sessions with at least 10% of retention.
“Sports” (20%) × “Entertainment” (30%).

Temporal Analysis

Life Span



Video life span - total and per category (CDF).

- ▶ The evolution of views follows a logarithmic function.
- ▶ Categories differs on the convergence rate.
- ▶ 77% of the views from “Sports” occur on their first day. While videos from “Science” achieve only 29% on their first day.

Conclusions

- ▶ We have presented an extensive analysis of video access patterns in Brazilian MSM portals.
- ▶ Our investigation revealed interesting static and temporal patterns that can be compared with results in UGC websites.
 - ▶ the two most popular categories in YouTube are among the least prevalent in MSM portals.
 - ▶ YouTube videos are relatively much shorter than MSM ones.
 - ▶ life span depends on category but, in general, it is very short.
- ▶ The access patterns can be used by MSM portals to improve service quality and enhance users' experience.
- ▶ **Future work:**
 - ▶ investigate social aspects of MSM portals.
 - ▶ synthetic workload generation.

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.