Machine Learning and Popularity Prediction of a Video Content

Nesrine Ben Hassine*†, Dana Marinca†, Pascale Minet*, Dominique Barth†

*Inria, Rocquencourt, 78153 Le Chesnay Cedex, France,

†DAVID, University of Versailles,

78 000 Versailles, France

Abstract— In Content Delivery Networks (CDNs), the knowledge of the popularity of video contents helps the manager to take efficient decisions about which video contents should be cached near the end users and also about the duplication degree of each video to satisfy the end user Quality of Experience. This paper focuses on predicting the popularity of video contents, in terms of number of solicitations. For that purpose, different software entities, called experts compute popularity values of video contents. Each expert uses its own prediction method. The expert prediction accuracy is evaluated by a loss function as the discrepancy between the prediction value and the real number of solicitations. The simulations based on real YouTube traces show that the accuracy of prediction is improved by splitting the video content profile in contiguous phases. Different prediction methods are compared and also different phase change-points detection methods are evaluated in order to identify the method (or method parameters) minimizing the cumulated discrepancy compared to real solicitations of video contents.

Keywords—CDN, popularity prediction, YouTube contens, phase profile

I. INTRODUCTION

A. Context

Due to the increasing use of the Internet services, there has been an important growth in network traffic. This results in decreases in service quality. To overcome the limitation of the Internet, Content Delivery Networks (CDNs) provide a method for a more efficient delivery of contents to end users. CDNs reduce the traffic directed to a single site via caching. They distribute contents geographically; heavily requested contents are cached at various locations, close to users who request them the most. Thereby, it is crucial to determine whether a content is requested sufficiently often so that it is worth caching it. The user request is redirected to the nearest location storing the solicited content. Hence, CDNs improve the network performance and fasten the delivery of requested contents to any end user location. Moreover, cache management is a challenging task. Given its finite storage capacity and the dynamicity of the demand, each stored content has to be continuously monitored so that it is pushed out from the cache once it is no longer required. These issues require an in-depth analysis of the dynamics of contents requests.

In the Global Internet phenomena report 2014 [1], streaming traffic remains the major source of Internet traffic in the world. Understanding the dynamics of the number of requests for streaming traffic is consequently both critical and indispensable for a better resource management. Streaming contents are mainly formed of video contents and they are basically originated from Video on Demand (VoD) platforms and User Generated Content (UGC) systems such as YouTube.

To better manage the enormous number of video contents, Youtube maintains various categories of video contents. The most popular video categories are cartoon, music, sports, movies, news... where some contents may belong to several categories. Hence, this classification helps people find the contents interesting them more easily.

The behavior of video contents over time reveals distinct patterns of popularity evolution. A huge number of contents are viewed only a few times. Others are largely required. There are contents that are popular for a long period of time whereas others are much solicited for few days or even hours. Hence, it is possible to estimate the future popularity of content by examining the types of popularity growth behavior that contents display over time. Therefore, demands can have different behaviors depending on the content and the time interval.

As a matter of fact, popularity evolution pattern is a series of phases. Each phase may be very different from the others, whereas it retains an homogeneous behavior inside. Let us define the concept of phase:

Definition 1: A phase can be defined as an interval of time during which a measured metric remains relatively stable.

Thus, if a phase is properly identified, the prediction within a phase should be more accurate. That is what we want to prove in this paper.

The paper is organized as follows. In Section I-B, we describe different methods to detect phase changes. In Section II, we give the problem statement and define the experts used to predict popularity of video contents. Results obtained from real traces are presented and explained in Section III, where we use different methods to detect phase changes. Finally, we conclude in Section IV and list some further work.

B. Related Work

Let us consider a time series of data characterizing a phenomenon monitored. Usually, it is possible to identify subsets of contiguous data where a same model can be applied

to represent the phenomenon considered. It is then of the utmost importance to identify the time points at which changes occur in the data collected.

A large number of change-points detection methods has been proposed to tackle this problem. Most of the studies consider a data set consisting of observations at discrete times $X_1..X_n$ spaced at uniform time intervals, let us assume we can partition the data into an unknown number K of partitions, $S_1..S_k$. One of the most difficult issues is estimating the number of change points.

Bai and Perron [3] propose a method with an a priori fixed number of change-point s. Let m be the maximum number of change-point s. For each $k \leq m$, a change-point location is the one ensuring the smallest within-segment square sums. To select the best number of change-point s, they pick the partition that minimizes the Bayesian Information Criterion (BIC). The criterion of selection could be SIC (Schwarz Information Criterion) or AIC (Akaike Information Criterion)

In [4], Taylor developed a change point detection method to detect changes in time series (CUmulative SUM). This approach is based on the mean-shift model written as $X_i = \bar{X} + \varepsilon_i$. $\bar{X}$ is the simple average and ε_i is the residual term associated with the ith observation. Residuals are assumed to be independent and identically distributed with mean zero. The cumulative sum is not the cumulative sum of the values but the sum of differences between the values and the average. The change-point is detected through searching for the maximum of absolute CUSUM. Once a change-point is detected, the time series is splitted into two segments and the procedure is repeated.

Structural Change Model (SCM) is another mean-shift model based method [4]. Initially, they split time series into two segments at any location k. Let the sum of squared residuals be defined as the mean square error estimator.

$$MSE(k) = \sum_{i=1}^k (X_i - \bar{X}_1)^2 + \sum_{i=k+1}^n (X_i - \bar{X}_2)^2$$

Where $\bar{X}_1 = \frac{\sum_{i=1}^k X_i}{k}$, $\bar{X}_2 = \frac{\sum_{i=k+1}^n X_i}{n-k}$. The change-point is found by minimizing the sum of squares residuals and located at $\hat{k} = argmin_{1 \leq k \leq N} MSE(k)$.

As CUSUM and SCM, Circular Binary Segmentation (CBS) [5][6] needs the whole data to perform its analysis. Each time a change-point is detected, data are splitted into two segments at the change-point location. The method stops when no change-point can be found in any segment. The location of change-point is the one maximizing the absolute value of the t-static, where t-static is a statistic test given by

$$T_{i,j} = \left(\bar{Y}_{i,j} - \bar{Z}_{i,j} \right) * s_{ij}^2$$

Where $\bar{Y}_{i,j}$ is the mean over $\{X_{i+1}..X_j\}$, $\bar{Z}_{i,j}$ is the mean over $\{X_1..X_{i+1}, X_{j+1}..X_n\}$ and s_{ij}^2 is the corresponding mean squared error.

Instead of estimating that a change-point has occurred, the Bayesian method computes the probability of a hypothesis. It expresses uncertainty about the number and location of change-point s. It computes the probability of a change-point at each location. In the Bayesian analysis, the posterior probability measures how likely we have a change-point at t. It is the weight given to hypothesis h where $h = \{there\ is\ a\ changepoint\ at\ t\}$. This probability is computed using the Bayes' theorem given by

$$P(h|E) = \frac{P(E|h) * P(h)}{P(E)}$$

where E is an evidence. Once we obtain the posterior probability, it is used to compute the posterior prediction distribution. This latter is a weighted average of the predictions of each individual hypothesis.

All the methods mentioned above are off-line and could be applied only once all the data are collected. In this paper, we compare an off-line method based on the Bayesian approach with two on-line methods that we will present in Section II-D.

II. THEORETICAL FRAMEWORK

A. *Problem Statement*

The goal of this paper is to predict the number of solicitations of individual video contents. We use real data sets of number of solicitations of video contents extracted from YouTube platform.

We adopt a machine learning approach and propose a model using several software entities called experts. Each expert computes and predicts future solicitation number for individual video contents using its own computation method (for example Basic, DES) with its own parameter set (e.g. size of its observation window, smoothing parameter).

We use a set of experts E, each expert $E_i \in E$ computes at time E the value $p_{i,t+1}$, representing the predicted number of solicitation of a given video, using its own prediction strategy. The goal of each expert is to predict a solicitation number as close as possible of the real solicitation number y_{t+1}. We will describe two prediction strategies in Section II-B. The accuracy of experts prediction is evaluated by the cumulative loss, as defined in Section II-C.

Focusing on the evolution of the number of solicitations of a video content, called video profile, we observe the existence of phases. A phase is characterized by a similar evolution of the number of solicitations as shows the video content profile depicted in Figure 1. The video contents considered belong to the Music and Cartoon categories. Figures on the right represent the number of solicitations as a function of time. For each profile, we represent also the cumulated number of solicitations (Figure 1). Tics represent phase changes.

The goal of this study is to answer the following questions:

- Can we improve the prediction accuracy through the cooperation of several experts for a given video content instead of one expert prediction?

- Does the use of phases improve the accuracy of predictions?

- Has the phases detection method an impact on the quality of predictions?

B. Experts

We recall that each expert prediction concerns an individual

video content. At a given time t, each expert $E_i \in E$ computes and predicts the value $p_{i,t+1}$, which will be compared to the real value at the time $t + 1$.

a) Music content: original profile. b) Music content: cumulative profile.

c) Cartoon content: original profile. d) Cartoon content: cumulative profile.
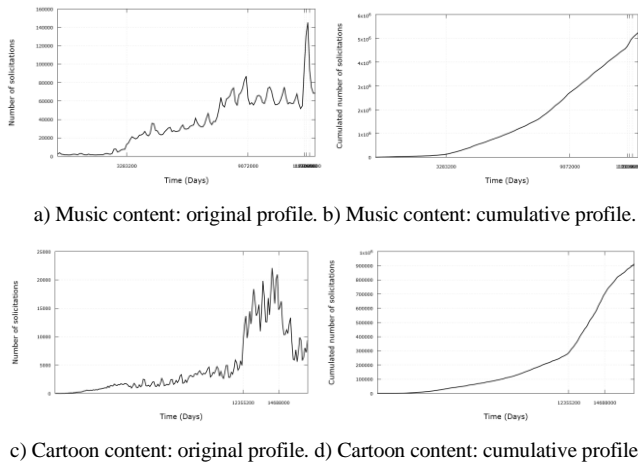
Fig. 1. Phases in the number of solicitations of a video content.

We define hereafter two concepts: (a) the observation window ow, a past time interval ending at time t, containing past consecutive values of the number of solicitations for the considered video and (b) the prediction window pw, starting at time t, representing the future time interval at the end of which the accuracy of the prediction $p_{i,t+1}$ will be evaluated.

1) DES expert:

DES experts, Double Exponential Smoothing, [2] have been introduced to cope with trends (e.g. increase) in the data to predict. A DES expert applies two exponential smoothing:

- first on the value y_t observed at time t when it computes $S'_{i,t}$. We have: $S'_{i,t} = \alpha y_t + (1 - \alpha)S'_{i,t-1}$, where α is the smoothing factor, with $0 < \alpha < 1$.

- second on the value of $S'_{i,t}$ when it computes $S''_{i,t}$, using the same smoothing factor α. We have: $S''_{i,t} = \alpha S'_{i,t} + (1 - \alpha)S''_{i,t-1}$, with $S'_{i,0} = S''_{i,0} = y_0$.

The value $p_{i,t+1}$ predicted by any Double Exponential Smoothing (DES) expert E_i at time $t + 1$ is given by: $p_{i,t+1} = L_{i,t} + T_{i,t}$, where $L_{i,t} = 2S'_{i,t} - S''_{i,t}$ denotes the estimated level and $T_{i,t} = \frac{\alpha}{1-\alpha}(S'_{i,t} - S''_{i,t})$ the estimated trend. A DES expert has two parameters: α the smoothing factor and ow the size of the observation window, which can be different for different experts.

2) Basic expert:

The basic expert is the expert that at time $t + 1$ predicts $p_{i,t+1} = y_t + (y_t - y_{t-1}) = 2y_t - y_{t-1}$. In other words, it predicts the last known value plus its increment with the previous value. Such an expert is very simple to implement

C. Best expert

Informally, the best expert is the one providing the most accurate predictions. This accuracy can be evaluated by the loss incurred by this expert.

Definition 2: For any expert E_i, we define its instantaneous loss at time t as $l_{i,t} = |y_t - p_{i,t}|$ and its cumulated loss at time t as $L_{i,t} = \sum_{k=0}^{t} l_{i,k}$.

Definition 3: For any set of experts E, we define the best expert as the expert $\in E$ that minimizes the cumulated loss. Notice that the expert that minimizes the cumulated loss also minimizes the average instantaneous loss. We now propose to extend the concept of best expert to the best expert on a given phase as follows:

Definition 4: For any given phase, for any set of experts E, we define the best expert for this phase as the expert $\in E$ that minimizes the cumulated loss on the phase.

D. Phase Detection

In this paper we distinguish several methods to detect phase changes.

- **Detection by a sudden increase of the instantaneous loss of the Basic expert**. This detection is simple to implement and can easily be done on-line.

- **Detection by applying Bayesian inference method.** A phase change is detected by applying Bayesian inference method, implemented in R tool. This detection requires the knowledge of the whole profile and consequently is done off-line only. We use the module implemented in the R tool.

- **Detection of a contiguous period**. Each phase has a constant duration of k time units, with k an integer greater than 1. This detection is very simple but may not correspond to a change in the profile of solicitations. Such detection is well adapted to profiles presenting periodic behaviors.

This paper studies the impact of phase detection method on the prediction accuracy.

E. The Prediction Method

We consider several video profiles belonging to different content categories. In an off-line analysis we run a lot of DES experts analyzing the considered profiles and predicting the number of solicitations. Among the identified experts we determine a subset of DES experts, having a cumulated loss less than 102% of the cumulated loss of the best DES expert on the entire profiles. These experts are characterized by particular values of the parameters α and ow. This simulation helps us to identify a short list of DES experts. Furthermore, we consider the profile segmentation according to the

Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
PEMWN - 2015 Conference Proceedings

different methods specified in Section II-D. For each change-point detection method, the cumulated loss of each expert is evaluated at the end of each phase. The best expert on each phase is the expert having the smallest cumulated loss. At the end of the profile, the prediction cumulated loss represents the sum over all phases of the cumulated loss of the best expert on each phase.

III. SIMULATION RESULTS

A. Determination of the Best DES Experts

We considered 50 video content solicitation profiles collected from YouTube plate-forme, belonging to different video categories (society, music, films, children, etc.). To determine the best values of the smoothing factor α and the observation window ow for the DES experts, we run many simulations on collected profiles. We succeeded to identify a list of six DES experts, the best ones, which are given in Table I. The parameter $owmax$ corresponds to the duration of the considered phase.

TABLE I. BEST DES EXPERTS.

α	$owmax$
0.8	$ow = 7$ ou $owmax$
0.93	$ow = 7$ ou $owmax$
0.9999	$ow = 7$ ou $owmax$

The next simulations are based only on the selected subset of the best experts presented in the previous Table.

B. Phase Change Detection

We now focus on the detection of phase changes. We apply the three methods described in Section II-D. For the detection of a contiguous period, we use 7 days, 14 days and 21 days.

C. Prediction Evaluation

1) No phase detection:

In the first series of experiments, we consider video contents profiles without distinguishing any phase. We test the Basic expert described in Section II.B.2 and the best DES experts given in Table I. Results obtained on real traces show that for different profiles the associated best experts could be different. In Figure 2, the video content considered belongs to the category Music. For this content, the expert minimizing the cumulated loss is the DES expert with $\alpha = 0.9999$ whatever the size of the observation window. The Basic expert has close results.
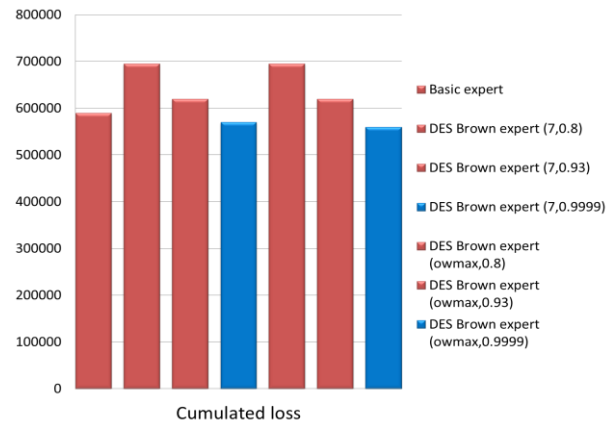

Fig. 2. The total loss of selected experts over the whole music contents

We now focus on a video content of category Cartoon. For this content, the Basic expert drastically outperforms the DES experts, as depicted in Figure 3. Notice that all the DES Brown experts have approximately the same cumulated loss. Thereby, the Basic expert has very good results in both cases.
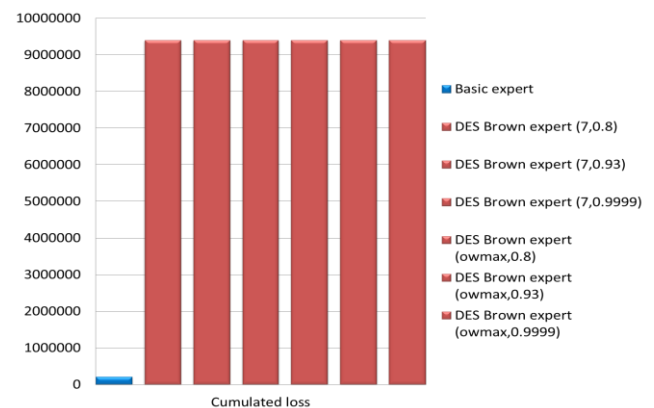

Fig. 3. The total loss of selected experts over the whole cartoon contents.

2) Detection of Phase Changes:

In the second series of experiments, we want to answer the following question: can we improve the accuracy of prediction by detecting phase changes? Tested video profiles are segmented using the methods given in Section II-D. Each method has its detected points which differ from the output of other methods. Results are depicted in Figure 4 for a cartoon video content.
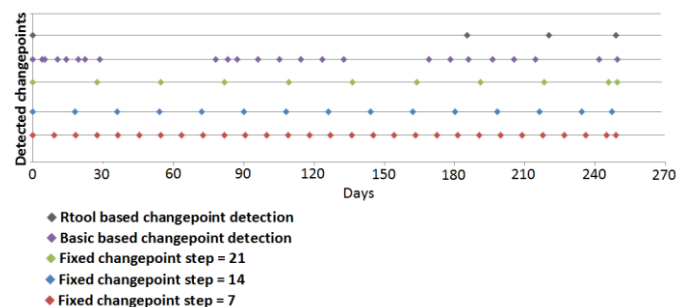

Fig. 4. Cartoon content: Detected change-points using different methods.

Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
PEMWN - 2015 Conference Proceedings

We will compare the cumulated loss obtained by these methods with the cumulated loss obtained when ignoring the profiles phases, denoted as "No change-point detection". More precisely, two cumulated losses are compared over each profile:

(1) Using "No change-point detection" method, we compute the cumulated loss of the best expert at the end of the profile;

(2) Using each change-point detection method, we compute the sum over all profile phases of the cumulated loss of the best expert per phase. At the end of each phase, the cumulated loss of each expert is reinitialized and another best expert can replace the previous one. As depicted in Figures 5 and 6, we observe that regardless of the method used, all change-point detection methods reduce the cumulated loss obtained without detecting phases. This can be explained by the expert changes at the end of each phase. These simulation results show that the change-point detection has a strong impact on the accuracy of the prediction.

As shown in Figures 5 and 6, Basic based change-point detection provides the largest cumulated loss compared to other change-point detection methods. Nevertheless, in both figures, the smallest cumulated loss is obtained when the change-point detection method uses a period of 7 consecutive days. Earlier studies [9] have shown that users' access follows a pattern over a week. This explains the results obtained when it is assumed that there exit a change every week. Based on this assumption, the cumulated loss could be further reduced if the end of the interval coincides with the weekend.

Thus, the best method to detect phase changes is the method considering fixed periods of 7 consecutive days. These results justify an approach based on detecting phase changes in video profiles evolution.

3) Behavior of experts near a phase change:
In the third series of experiments, we evaluate the behavior of experts near phase changes. For instance, in Figure 7, we compare the behavior of Basic and DES experts with and without change-point detection. We assume that we detect a change-point every period of 7 consecutive days. The prediction of experts with change-point detection deviates slightly from the real value. However, experts without change-point detection make large errors at the phase change, especially the Basic expert. We also observe that within a phase all selected experts provide predictions close to the real value.
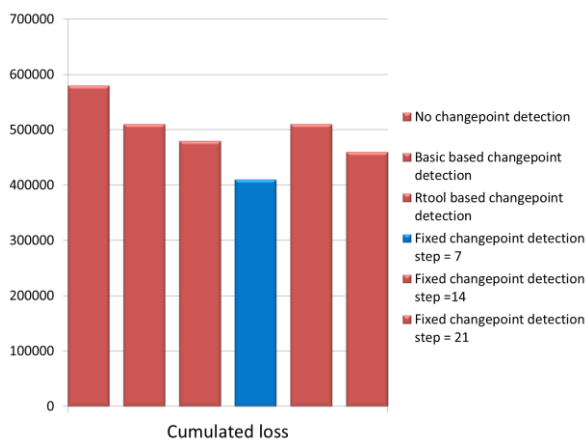

Fig. 5.   Music content: The minimum cumulated loss with no change-point detection vs the cumulated loss with different change-point detection methods.

Fig. 7.   Cartoon content: Prediction based on change-point detection using fixed period of 7 days.

IV.   CONCLUSION

Performances of Content Delivery Networks are usually measured by the hit ratio of user requests. The main goal of ny cache management policy is to maximize the hit ratio over CDN. In this paper, we propose to predict the popularity of video contents, in terms of the number of solicitations. This popularity can be used to decide which contents to be cached. The best expert predicting the number of solicitations is identified as the expert minimizing the cumulated loss over the profile. This expert provides the prediction the closest to the real value. To improve the accuracy of predictions, we defined different phases over each profile. The best expert per phase outperforms the best expert on the whole video content profile. However a false detection of a phase change (false positive detection) may decrease the quality of the prediction. Our further work will consist in designing a method detecting online phase changes without false positive detections.
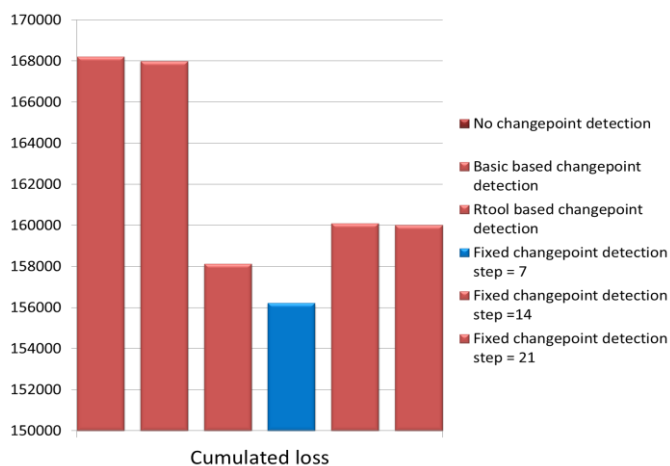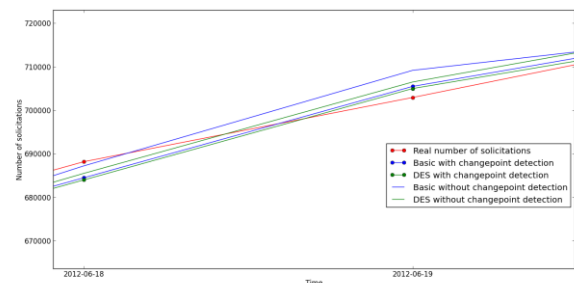

Fig. 6.   Cartoon content: The minimum cumulated loss with no change-point detection vs the cumulated loss with different change-point detection methods.

Special Issue - 2016

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
PEMWN - 2015 Conference Proceedings

REFERENCES

[1] Global Internet phenomena report, Sandvine, Tech. Rep., 2014. https://www.sandvine.com/downloads/general/global-internetphenomena/2014/1h-2014-global-internet-phenomena-report.pdf

[2] N. Bianchides, G. Lugosi, Prediction, Learning, and Games, Cambridge University, 2006

[3] J. Bai1, P. Perron, Computation and analysis of multiple structural change models, 2003.

[4] W A. Taylor, Change-point analysis: A powerful tool for detecting changes, 2000.

[5] Adam B. Olshen, E. S. Venkatraman, Circular binary segmentation for the analysis of array-based DNA copy number data, 2004.

[6] ES. Venkatraman, AB. Olshen A faster circular binary segmentation algorithm for the analysis of array CGH data, 2006

[7] K P. Murphy Machine learning: a probabilistic perspective, 2013.

[8] N. Ben Hassine, D. Marinca, P. Minet, D. Barth, Popularity Prediction in Content Delivery Networks, PIMRC 2015, Hong-Kong, China, September 2015.

[9] Y. Hongliang, Z. Dongdong, Ben Y. Zhao, W.Zheng, Understanding user behavior in large-scale video-on-demand systems, EuroSys 2006, New York, USA