International Journal of u- and e- Service, Science and Technology
Vol.9, No. 6 (2016), pp.433-444
http://dx.doi.org/10.14257/ijunesst.2016.9.6.39

The View Growth Pattern of User Generated Videos on YouTube

Renjie Zhou*, Dongchen Xia , Yuyu Yin, Jilin Zhang and Wei Zhang

College of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, China
Key Laboratory Complex System Modeling and Simulation, Ministry of Education, Hangzhou, China.
**rjzhou@hdu.edu.cn*

Abstract

With the rapid development of social media, video sharing sites like YouTube are getting more and more attention. Discovering the view growth pattern have become interesting topics for researchers as well as advertisers, media companies. In this paper, we analyze two aspects about video view growth pattern of YouTube videos. Firstly, the pattern of aggregated view is studied. It is found that the aggregated view rate peaks in the first few days, and falls quickly in the following days, and then decrease slowly during the consecutive weeks. Finally, the view rate tends to be a constant on the long run. The aggregated view count after a period of two months can be fitted with a linear line. Secondly, the view growth pattern of individual video is explored. The results indicate that the majority of videos peak at the very beginning of videos' lifetime, and the category of view sources causes the peak is different. The view count of individual video and the view count from each source item also stabilize after a period of two months, and we finally show the referring time and active period of each source item.

Keywords: *aggregated view pattern; individual view pattern; view rate; view source; stabilization of view count*

1. Introduction

Over the past few years, social media has played a more and more important role in our lives, and YouTube is one of the typical examples. As a video sharing site, YouTube has more than one billion registered users, and billions of videos are watched each day [1]. The development of social media has boosted the social marketing. As a result, through the data information on the YouTube to tap the potential of the law and valuable information has been the focus of various communities such as researchers, advertisers, and publishers.

In our previous work, we analyzed the main sources of video views, including the search engines and recommendation systems of YouTube, mobile terminals, channel subscriptions, embedding capability on web pages, blogs, and social networks [2]. However, there are still some interesting questions need to be answered. First, whether the aggregated view pattern of view count and view rate has distinct characteristics? Second, how different sources of views contribute the view growth of videos over time? Third, when individual video's popularity peak ? What category of view sources caused the peak? Fourth, whether view count of individual video, and the view count from each source item also stabilizes after a period of time? Fifth, what is the referring time and active period of each source item.

*Corresponding Author.

ISSN: 2005-4246 IJUNESST

In order to answer the questions raised above, we collected seven months of data for more than 6000 videos from YouTube. We extract the information and analyze it in the following two aspects. On the one hand, the aggregated view pattern is studied. One the other hand, the statistics of individual video view pattern is explored. Through the analysis of a variety of viewing patterns in these two aspects, we present the answers for the aforementioned questions.

The rest of this paper is organized as follows. In section 2, we survey the literature of existing related works. The dataset of YouTube videos is described in section 3. Section 4 analyzes the aggregated view pattern. The statistics of individual video view pattern is explored in section 5. Finally, section 6 concludes the paper.

2. Related Work

As the representative of the social media in recent years, YouTube has been gradually changing people's way of life, a lot of research works have been done on YouTube. Cha *et al.* conducted a broad analysis about the characteristics of video popularity on YouTube, such as the video age, the level of content aliasing or of illegal content [3-4]. Chatzopoulou *et al.* analyzed popularity in a comprehensive fashion by looking at properties and patterns in time and considering various popularity metrics, they found that view count is highly correlated with the number of comments, ratings, and favorites [5]. Figueiredo *et al.* characterized popularity growth for three different types of videos, namely, Quality, Viral and Junk video, show that popularity growth patterns depend on the video dataset [6]. Moreover, they presented a model of predicting popularity trends complementary to hits in the following work [7]. In their most recent work, the authors discuss the factors that impact the popularity dynamics of social media [8]. Chen *et al.* study how the video popularity changes over its lifetime for different types of videos, and proposed a lifetime model of online video popularity [9].

The works mentioned above are based on the analysis of the popularity from all the video sources. However, in this paper, we explore how each video source affects the overall viewing distribution and individual video's popularity.

In addition, a lot of popularity prediction models are proposed for online content including images, YouTube videos and Digg stories [1], [10-14]. Especially in [13], Szabo and Huberman compared the accuracy of using early popularity to predict the future popularity on Digg and YouTube. The results show that predictions on YouTube are harder than on Digg. Our work is a supplement to theirs. By classifying the view sources, we study the contribution of each video source in each time of the video. The results of this work can provide some useful hints for building better prediction model.

3. Data Set

The data set consists of 6099 videos, which were retrieved from the most recent uploaded videos through YouTube Data API on September 9th, 2010. Among the videos, 4309 were uploaded on September 8th and 1760 were uploaded on September 9th and 30 were uploaded on September 7th. After retrieving the videos, the metadata, statistics & data, and related video list of them were crawled every three days over time until April 10th, 2011. However, there were a few consecutive snapshots that the time period is not three days due to interrupts occurred during the crawling. The 17th and 18th snapshots have a time interval of 5 days and the 28th and 29th snapshots have a time interval of 6 days and the 65th and 66th snapshots have an interval of 5 days.

4. Aggregated View Pattern

4.1. Aggregated View Pattern of Videos

We first investigated how the daily aggregated view rate changes over time, which is computed as the view count accumulated between every two consecutive snapshots over the days between the two consecutive snapshots. As shown in Figure 1, the daily aggregated view rate fell down a lot from the first point to the second point and then decrease more slowly during the consecutive weeks. Finally, the view rate tends to be a constant on the long run. Figure 2 shows the accumulated view count over time. As can be seen from the figure, the growth of the accumulated view count is fitted with a power law curve very well.

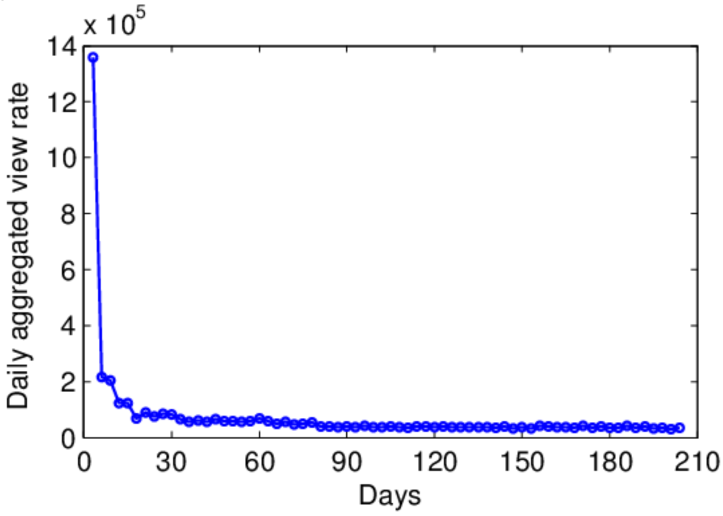


Figure 1. The Daily Aggregated View Rate of All Videos Over Time.

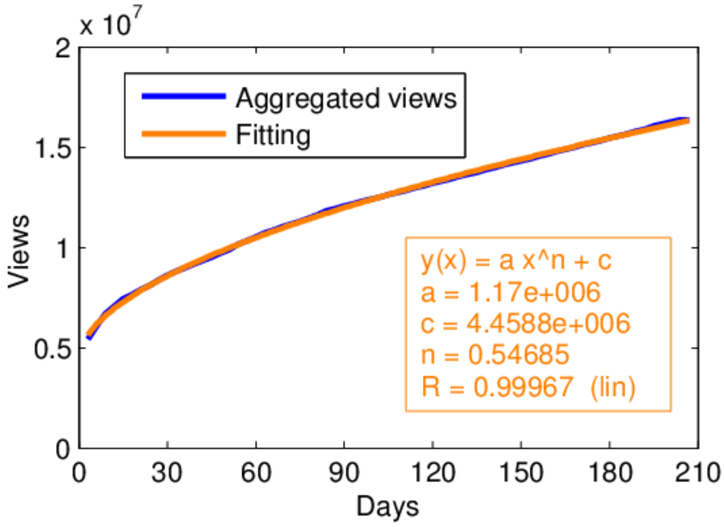


Figure 2. The Aggregated View Count of All Videos Over Time.

The aggregated view count after the 60th day can also be well fitted with a linear line using the least squares method, and the fit is evaluated using the goodness of fit R^2:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (1)$$

where y_i is the observed view at day i, $\hat{y}_i$ is the view from the fitted line at day i, and $\bar{y}$ is the mean of the observed view. The value of R^2 ranges from 0 to 1, where R^2 = 1 indicates the best fit. The goodness of fit for the aggregated view count after the 60th day is 0.998.

4.2. Aggregated View Pattern of Each Category of Sources

We then explore how different sources of views contribute the view growth of videos over time. We calculated the percentage of daily aggregated view rate contributed by each view source, which indicates the importance of each view source during each time interval. As can be seen in Figure 3, the percentage of view rate contributed by Search and Mobile is quite stable over time. Furthermore, the percentage of view rate contributed by Related is very low at the beginning; however, it quickly rises to approximately 50%. On the contrary, view rate contributed by Embed and AFSY (Ad+Featured+Subscribe+YouTube) is quite high at the very beginning and then drops over time. For more information of the view source category, we would like to refer you to our previous work [2].

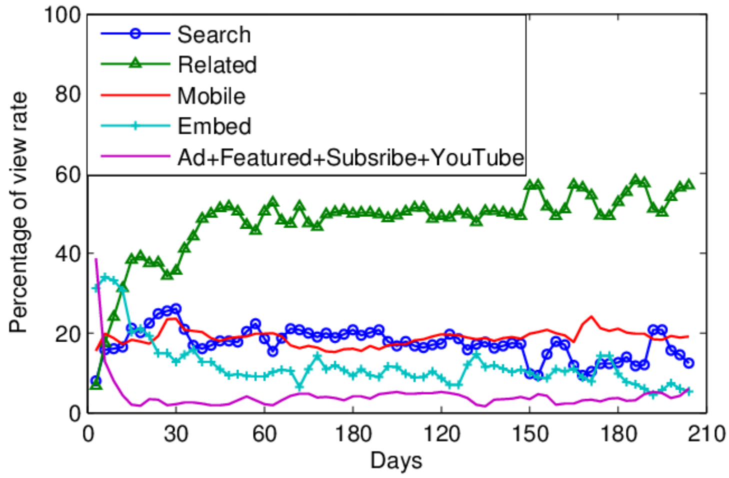


Figure 3. The Percentage of View Rate from Each View Source

In addition to the percentage of view rate from each source, we also examined the percentage of accumulated view count from each source.

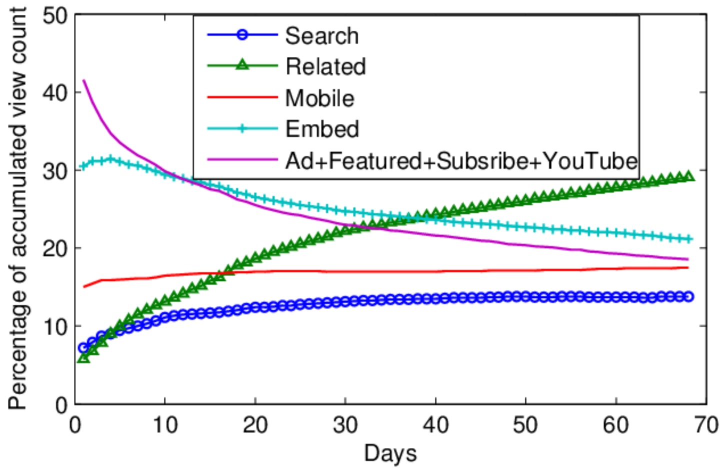


Figure 4. The Percentage of Accumulated View Count from Each View Source

As can be seen from the figure, the percentage of accumulated view count from both Search and Related increases as the time passes. More specifically, the percentage of accumulated view count from Search increases from 7.2% to 13.8%, and the percentage from Related increases from 5.7% to 29%. The percentage from Mobile keeps stable around 15%. However, the percentage from Embed and AFSY decreases quickly as the videos become older.

5. Statistics of Individual Video View Pattern

In the previous section, we found that the aggregated view count peaks at the very beginning of videos' lifetime. In this section, we will investigate whether this is a collective behavior or just caused by a few exceptional videos. Is there any common characteristic shared by videos that peak at the beginning? What is the difference between videos that peak at the beginning and those peak later? Whether view count of individual video, and the view count from each source item also stabilizes after a period of time, *e.g.* two months.

5.1. When Individual Videos Peak

In Figure 1, it is shown that the aggregated view rate peaked at the first few days. In this section, we examine whether the peak is caused by a few videos or the majority videos. We investigated the time of each video's view peaked and found that the majority of videos peaked at the first few days. As can be seen from Figure 5, approximately 36.8% and 31.5% of the videos are peaked at the first four days and consecutive three days, respectively. It means approximately 70% of the videos peaked in the first week. Other videos are almost peaked evenly across the rest of time.

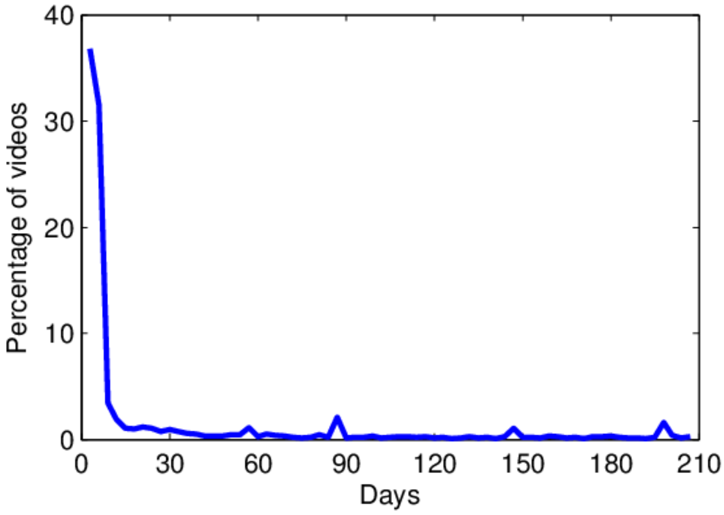


Figure 5. The Distribution of Peak Time

5.2. What Category of View Sources Cause the Peak

In this section, we investigate whether the peaks happened at different time were caused by different category of view sources. For each video, we examined when the view is peaked and identified the category of view sources that contributed the largest percentage of views during the peak time. We found that the early peaks are usually caused by Embed, YouTube Search, and YouTube Highlight *etc.* However, the later peaks are usually caused by Related, YouTube Search and Embed.

Table 1. Percentage of Peaks Caused by Different Categories

Time / Category	[0,10]	[0,30]	[30,210]
Embed	34.1	34.3	19.8
YouTube	14.0	13.0	6.9
Related	9.5	10.7	37.4
Search	26.1	26.5	24.0
Subscriber	9.0	8.4	0
Mobile	7.3	7.1	11.8

5.3. Stabilization of Individual Video View

It is known that the aggregated view stabilizes after a period of about two months. In this section, we investigate when individual videos stabilize. In this paper, stabilization of individual video view means either the view count stop growing or grows linearly. The stability of video views is measured by stability coefficient S, which is defined as $S = 1$ if view count stop growing, otherwise the goodness-of-fit of linear regression $S = R^2$. For a video, whose upload time was $t = 0$, if there was a time point T ($T < 210$ days) since then the stability coefficient of the video was high enough, then we say it took time T for the video to stabilize. The distribution of stabilization time of videos is shown in Figure 6. For example, we take the stability coefficient 0.95 as the standard to judge how long it took for a video to stabilize, 22.4% of the videos had been stabilized just after it was uploaded and approximately 60% of the videos were stabilized within two months. On the other hand, 20.4% of the videos need more than five months to stabilize and among them 5.9% of the videos had not yet been stabilized within the crawling period. The figure also shows the distribution of stabilization time for a standard of stability coefficient 0.90, under which more than 75% of the videos were stabilized within two months.

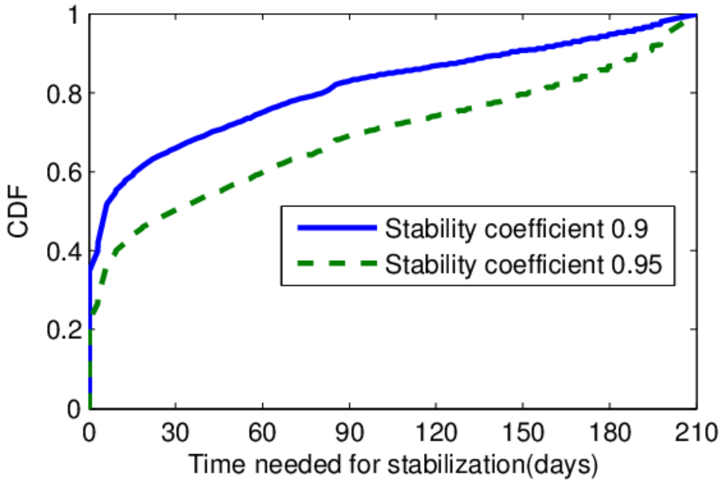


Figure 6. The Distribution of the Time Needed for Stabilization

5.4. Stabilization of View from Each Source Item

In this section, we investigate whether the view from each source item is also stabilized after a period of two months. We measure the stability of view from each source item with the stability coefficient defined before. The result is shown in the third row of table 2, which tells that the stability coefficient of view from each source item is about 0.88. So far, we have learned that all the three levels have a high stability coefficient after a period of two months, with the aggregated view count has the highest stability coefficient. Table 3 shows the average stability coefficient for source items from each category.

Table 2. The Stability Coefficients

Time / Type	[60,210]
Aggregated Videos S	0.998
Video-level Average S	0.876
Item-level Average S	0.882

Table 3. The Stability Coefficients for Each Category

Time / Type	[60,210]
Embed	0.876
YouTube	0.861
Related	0.847
Search	0.910
Subscribe	0.949
Mobile	0.850

5.5. Referring Time and Active Period of Each Source Item

In this section, we investigate when a source item referred to a video and how long it lasts for leading views to the video. The reason we investigate this problem is that we want to know what factors determine the stabilized rate. If some categories show up very early but last for a very short time and some categories show up much later but last for much longer. Then we can say the later categories determine the stabilized rate. The referring time of a source item to a video is represented by the age of the video when it is referred. The active days of a source item is represented by the number of days during when the view count from the source item continues growing. Since we have a limited period of data, the maximum active period for a source item is 210 days. The average of referring time and active days is shown in Table 4 and the distribution of referring time and active days is shown in Figure 7 and 8, respectively. In table 4, column count represents the number of source items for each given category. As shown in the table, Related and YouTube Search have the two largest number of source items, while featured and ad only have a very small number of source items. The average referring time of source items from subscriber category is only 0.6, which means the first view from a subscriber of a video is always within one day. We also find that the source items from Related are relatively later than items from Embed, Featured, YouTube etc, however, the difference is not that significant. For the active period, it is understandable that source items from mobile category has a large number of active days, since it actually represents the client type not the accessed place of a video. However, it is unexpected that the source items from featured category have the largest average active days. The source items from Subscriber, Google Search and YouTube have a relative short active period. The source items from Related and Embed have a moderate active period. As shown in Figure 7, the majority of source items from categories of Subscribe, Embed, YouTube pages and referral from are from the first day of a video been uploaded. While sources items from Related and Google Search take a while to refer to a video.

Table 4. The Referring Time and Active Period of Source Items

Type	Count	Referring time	Active days
Embed	12732	11.7	51.9
Youtube	9308	11.0	31.4
Related	20987	24.8	77.9
Search	29430	14.6	42.1
Subscribe	1694	0.6	20.3
Mobile	4182	11.5	127.3

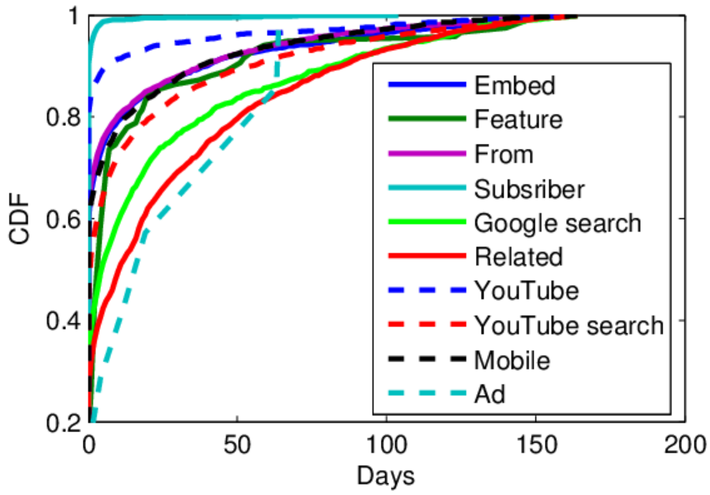


Figure 7. The Distribution of Referring Time

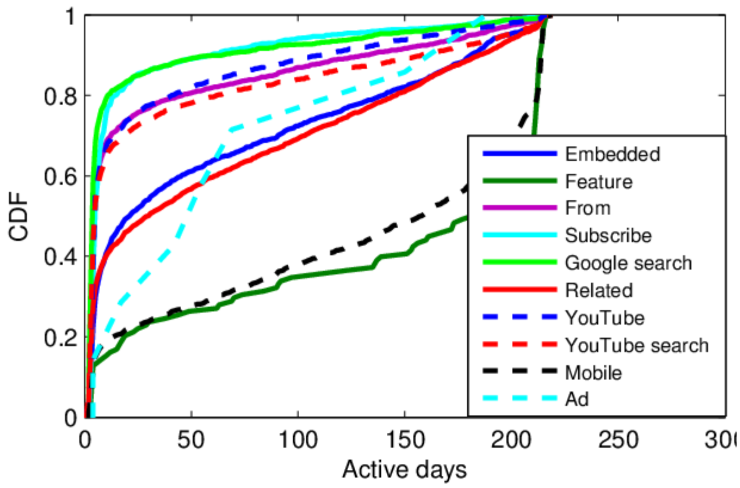


Figure 8. The Distribution of Active Days

In summary, it is shown that Related shows up later but lasts for a much longer period than other type of view source. On the contrary, Subscribe, YouTube Highlight and Embed show up earlier and have a smaller number of active days.

6. Conclusion

In this paper, we investigated the video view growth pattern by collecting the video information on the YouTube sharing website. The analysis is divided into two parts. In first part, the aggregated view pattern is studied; in the second part, the statistics of individual video view pattern is explored. Through the analysis of two aspects, we have arrived at the following conclusions. Firstly, the daily aggregated view rate drops very quickly during the first week, and then decrease more slowly during the consecutive weeks. Finally, the view rate tends to be a constant on the long run. This is also agreed with our result that the aggregated view count of all of the new videos grows linearly since 60th day. Moreover, the view count of individual video and the view count from each source item also stabilize after a period of two months. Secondly, the phenomenon that the aggregated view rate peaks at the very beginning of videos' lifetime is caused by the composition of majority of videos. Furthermore, the early peaks are usually caused by

Embed, YouTube Search, and YouTube Highlight *etc.* However, for the rest of videos, their peak view rates show up in the later period and are usually caused by Related and YouTube Search. Thirdly, the search engines and recommendation systems are the two most important view sources persistently driving views for videos in the long run, as the percentage of accumulated view count from search increases from 7.2% to 13.8%, and the percentage from Related increases from 5.7% to 29% from the upload time to the end time of the investigation. We also demonstrated that Related shows up later but lasts for a much longer period than other type of view source, while, Subscribe, YouTube Highlight and Embed show up earlier and have a smaller number of active days.

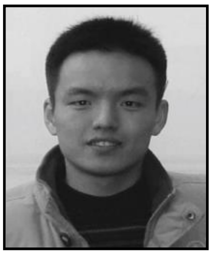
Acknowledgment

The authors are grateful to the anonymous reviewers for their valuable comments and to the editors for their work that improved this paper. This work was supported by NSF of Zhejiang under grant NO.LQ13F020017, LY14F020044, LY16F020018, and NSF of China under grant NO.61300211, 61572163, 61472112, J1524009 and National Key Technology Research and Development Program of China under grant No.2014BAK14B04.

References

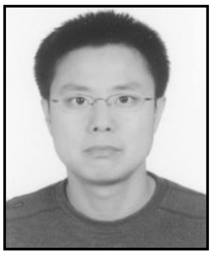
[1] M. Ahmed, S. Spagna, F. Huici and S. Niccolini, "A peek into the future: predicting the evolution of popularity in user generated content", Proceedings of the sixth ACM international conference on Web search and data mining, **(2013)**, pp. 607–616.
[2] R. Zhou, S. Khemmarat and L. Gao, "The impact of youtube recommendation system on video views", Proceedings of the 10th ACM SIGCOMM conference on Internet measurement, **(2010)**, pp. 404–410.
[3] M. Cha, H. Kwak, P. Rodriguez, Y. Ahn and S. Moon, "I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System", Proceedings of the 7th ACM SIGCOMM conference on Internet measurement. ACM, **(2007)**, pp. 1-14.
[4] M. Cha, H. Kwak, P. Rodriguez, Y. Ahn and S. Moon,"Analyzing the video popularity characteristics of large-scale user generated content systems", IEEE/ACM Trans Networking (TON), vol. 17, no. 5, **(2009)**, pp. 1357–1370.
[5] G. Chatzopoulou, C. Sheng and M. Faloutsos, "AFirst Step Towards Understanding Popularity in YouTube", INFOCOM IEEE Conference on Computer Communications Workshops, **(2010)**, pp. 1–6.
[6] F. Figueiredo, F. Benevenuto and J. M. Almeida, "The tube over time: characterizing popularity growth of youtube videos", Proceedings of the fourth ACM international conference on Web search and data mining, **(2011)**, pp. 745–754.
[7] F. Figueiredo, "On the prediction of popularity of trends and hits for user generated videos", Proceedings of the sixth ACM international conference on Web search and data mining, **(2013)**, pp. 741–746.
[8] F. Figueiredo, J. M. Almeida, A. M. A. Gonc and F. Benevenuto, "On the dynamics of social media popularity: A youtube case study", ACM Trans Internet Technol (TOIT), vol. 14, no. 4, **(2014)**, pp. 24.
[9] L. Chen, Y. Zhou and D.M. Chiu, "A lifetime model of online video popularity", 2014 23rd International Conference on Computer Communication and Networks (ICCCN), **(2014)**, pp. 1–8.
[10] K. Lerman and A. Galstyan, "Analysis of social voting patterns on dig", Proceedings of the first workshop on Online social networks, **(2008)**, pp. 7–12.
[11] K. Lerman and T. Hogg, "Using a model of social dynamics to predict popularity of news". Proceedings of the 19th international conference on World wide web, **(2010)**, pp. 621–630.
[12] H. Pinto, J. M. Almeida and A. M. A. Gonc¸ "Using early view patterns to predict the popularity of youtube videos", Proceedings of the sixth ACM international conference on Web search and data mining, **(2013)**, pp. 365–374.
[13] G. Szabo and B. Huberman, "Predicting the popularity of online content", Commun ACM, vol. 53, no. 8, **(2010)**, pp. 80–88.
[14] J. Yu, Y. Rui and D. Tao, "Click prediction for web image reranking using multimodal sparse coding", IEEE Trans Image Process(TIP), vol. 23, no. 5, **(2014)**, pp. 2019–2032.

Authors



Renjie Zhou, He is an assistant professor in School of Computer Science and Technology at Hangzhou Dianzi University, Hangzhou, China. He received his Ph.D. degree from Harbin Engineering University, Harbin, China, in 2012. He was a visiting scholar in the Department of Electrical and Computer Engineering at the University of Massachusetts at Amherst. Currently, he is working at the cloud computing research institute of Hangzhou Dianzi University. His research interests include measurement and analysis of online social networks, and network security.

Dongchen Xia, He is a Master candidate in School of Computer Science and Technology at Hangzhou Dianzi University, Hangzhou, China. He is with the cloud computing research institute of Hangzhou Dianzi University. Before joining Hangzhou Dianzi University, he was a college student in Zhejiang Shuren University from 2010 to 2014. His current research areas is measurement and analysis of online social networks.



Yuyu Yin received the Ph.D. degree in computer science from Zhejiang University, Zhejiang, China, in 2010. He is currently an Associate Professor with Hangzhou Dianzi University, Hangzhou, China. His research interests include service computing, cloud computing, and middleware techniques.

Jilin Zhang received the PhD degree in Computer Applied Technology from University of Science Technology Beijing, Beijing, China, in 2009. He is currently an associate professor at Hangzhou Dianzi University, China. His research interests include High Performance Computing and Cloud Computing.

Wei Zhang is an associate professor in School of Computer Science and Technology at Hangzhou Dianzi University, Hangzhou, China. He received the BE degree in School ofInformation Science and Engineering of Wuhan University of Science and Technology in China in 2000, and he received the MEc and PhD degree in Computer School of Wuhan University in China in 2004 and 2008, respectively. His research interests include network application and Intelligent Computing.

