Exploring the User-Generated Content (UGC) Uploading Behavior on YouTube

Jaimie Y. Park
Division of Web Science and Technology, KAIST
Daejeon, Korea
jaimie@islab.kaist.ac.kr

Jiyeon Jang
Department of Computer Science, KAIST
Daejeon, Korea
jiyjang@kaist.ac.kr

Alejandro Jaimes*
Yahoo Labs
New York City, USA
ajaimes@yahoo-inc.com

Chin-Wan Chung
Dept of Computer Science
Div of Web Science and Tech, KAIST
Daejeon, Korea
chungcw@kaist.edu

Sung-Hyon Myaeng
Dept of Computer Science
Div of Web Science and Tech, KAIST
Daejeon, Korea
myaeng@kaist.ac.kr

ABSTRACT

YouTube is the world's largest video sharing platform where both professional and non-professional users participate in creating, uploading, and viewing content. In this work, we analyze content in the music category created by the non-professionals, which we refer to as user-generated content (UGCs). Non-professional users frequently upload content (UGCs) that are parodies, remakes, or covers of the music videos uploaded by professionals, namely the official record labels. Along with the success of official music videos on YouTube, we find the increased participation of users in creating the UGCs related to the music videos. In this study, we characterize the UGC uploading behavior in terms of what, where, and when. Furthermore, we measure the relationship between the popularity of the original content and creation of the related UGCs. We find that the UGC uploading behavior is different depending on the types of the UGC and across different genres of music videos. We also find that UGC sharing is a highly global activity; popular UGCs are created from all over the world despite the fact that the popular music videos originate from a very limited number of locations. Our findings imply that utilizing the information on re-created UGCs is important in order to understand and to predict the popularity of the original content.

Categories and Subject Descriptors

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

*This work was performed while the author was a visiting professor at KAIST, Division of Web Science and Technology, under the WCU (World Class University) program.

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2576945.

General Terms

Human Factors, Measurement

Keywords

User Generated Content; YouTube; Popularity Analysis; Digital Media; Social Media; Video Streaming Service

1. INTRODUCTION

YouTube is the largest video sharing site in the world. It has been reported that more than 1 billion unique users visit the site, watching over 6 billion hours of videos on a daily basis[1]. YouTube is unique in that it is a truly participatory platform, where many users are self-publishing consumers [4]. In other words, a large portion of YouTube content is generated by non-professionals. We refer to such content as user-generated content (UGCs). With the advancement of video recording and editing technologies, the quality of UGCs has become more compelling to the audience. In the early days of the YouTube service, there was a perception that user-generated content was inferior to professionally produced content[2]. And to this day, professional content still commands higher ad rates. However, we have witnessed a continuous growth in the extent to which users participate in creating and viewing UGCs, which highlights the importance of understanding the phenomenon at a deeper level.

In this study, we put our focus on the videos that belong to the "music" category. We are particularly interested in observing the phenomenon in which users create parodies, remakes, or covers of the music videos uploaded on YouTube. When a music video becomes a hit on YouTube, we see many related UGCs springing up on YouTube as well. For instance, when a Korean pop "Gangnam Style" by the artist Psy became a hit in 2012, millions of parodies, re-

[1]YouTube statistics. retrieved on Feb 2nd, 2014. http://www.youtube.com/yt/press/statistics.html.
[2]YouTube: growth in user-generated content is outpacing official videos. retrieved on Feb 2nd, 2014. http://www.billboard.com/biz/articles/news/1178328/youtube-growth-in-user-generated-content-is-outpacing-official-videos

makes, and covers flourished on YouTube. Inspired by this phenomenon, we study whether the popularity of a music video on YouTube is related to the volume of related UGCs. Furthermore, we explore what types of UGCs are being uploaded on YouTube, where, and when.

To the best of our knowledge, this is the first work that examines the phenomenon of creating music-related UGCs with respect to the success of its original video. The questions addressed in this study include, but are not limited to, the following:

- What types of music video-related UGCs do users create?
- Do people create different types of UGCs depending on the genre of the music?
- Which countries are involved in uploading UGCs? How geographically diverse is the UGC uploading behavior?
- When are UGCs uploaded? How long does it take for UGCs to be uploaded after the release of original music videos?
- Is there a difference in the UGCs upload time depending on the type (parodies, remakes, covers, etc.) of the UGCs?
- Is UGC uploading related to the popularity of music videos on YouTube?

The results of our analysis show that the music video popularity and UGC upload frequency are strongly correlated, highlighting the importance of analyzing the UGC uploading behavior in understanding the popularity of the original content. We further find that UGC uploading behavior is different depending on the types of the UGC and across different genres of music videos. We also find that UGC sharing is a highly global activity; popular UGCs is created all over the world despite the fact that the popular music videos originate from a very limited number of locations.

2. RELATED WORK

Studies on user-generated content online mainly fall into one of two bodies of research: characterizing the popularity of videos and predicting the popularity of videos. Cha et al.'s work [4] was one of the first works to collect and analyze large-scale YouTube data in order to understand the popularity lifecycle of videos. Following their work, there have been ongoing interests in investigating the temporal dynamics and evolution of video popularity among the research community as well as industry. Related studies mainly focus on modeling, classifying, and predicting the popularity of videos on YouTube [2, 5, 7, 8, 10]. Brodersen et al. [3] address the problem of analyzing the spatial properties in addition to the temporal properties of the videos. Baek and Shin [1] examine the relationship between cultural distance and intercultural consumption of content on YouTube. Susarla et al. [9] study the impact of social influence in popularity of content on YouTube. Our investigation is different from the existing body of work since we study the particular phenomenon where non-professional users "re"create the works of professionals. Although extensive efforts have been put into analyzing the popularity of videos on YouTube, there has not yet been an attempt to understand the behavior of recreating the content in forms of parodies, remakes, or covers. Since these recreated content can potentially influence the popularity of the original content, it is worthwhile to study this new facet of YouTube video popularity and sharing phenomenon.

3. THE YOUTUBE DATA

In this section, we describe the data used for analysis, and the method to collect the data.

3.1 Top Music Videos (MVs)

We first collected the 1,000 most-viewed YouTube videos using the YouTube API[3]. The YouTube API returns up to 1,000 most viewed videos across all categories available on YouTube. After the collection we selected the videos that belong to the music category. In total 461 out of the 1,000 most-viewed videos belong to the music category. The majority of them are official videos uploaded by the entertainment agencies or record labels, for many agencies manage accounts on YouTube in order to share content, collect opinions from the audience, and connect with fans. Official videos are easily distinguishable from the non-official videos, namely UGCs, because the former are tagged as premium videos by the YouTube service. The official videos were mainly newly released music videos, while some of them were clips of live performances or concert scenes. We chose the top 100 official "music videos" uploaded by the official agencies.

3.2 Top MV-Related UGCs

We further collected the user-generated content that is related to the top music videos. Before collecting the UGCs, we manually parsed out the title of the song and the name of the artist from the top 100 music video titles. We then used the YouTube API to search for and retrieve the UGCs whose titles contain the song title, artist name, and any of the following designated terms: *acoustic*, *cover*, *dance*, *flash mob*, *parody*, *reaction*, *remake*, and *remix*. These terms, which represent the popular types of UGCs being frequently uploaded by YouTube users, were selected through the following procedure: we collected the top 1,000 most viewed music-related UGCs uploaded by non-professionals, computed the term frequency for every word that appears in at least one of the 1,000 UGC titles, then picked the terms that refer to the type of UGCs. Figure 1 is an illustration

[3]YouTube API. https://developers.google.com/youtube/

Figure 1: Title Term Frequency Word Cloud for Most-Viewed UGCs

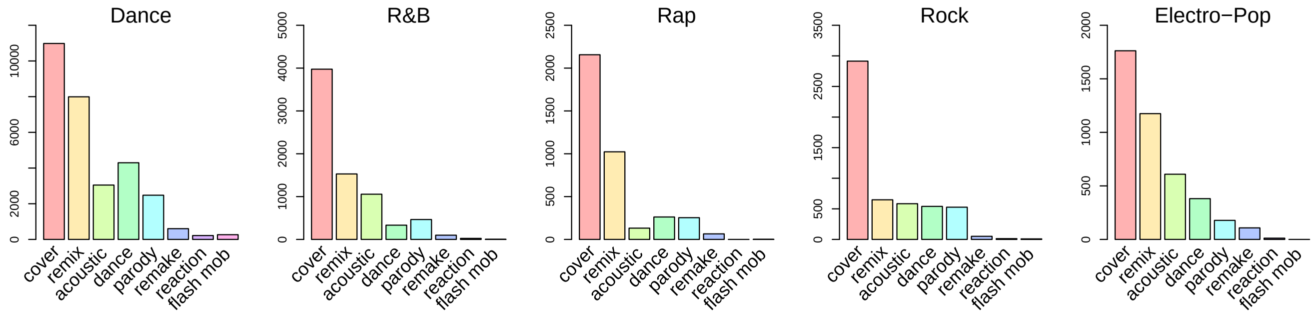


Figure 2: UGC Type Distribution by Music Genre

of the title term frequency in the form of tag cloud. Dance, video, official, cover, music, and live appear to be some of the most frequently used terms in the title to describe the most popular UGCs. We also find the terms parody, cover, live, acoustic, flash mob, and concert often being part of the video titles. Using the YouTube API, we retrieved the UGCs whose titles are composed of a <song title>, <artist name>, and <designated term>. For example, the top most viewed music video on YouTube is Psy's "Gangnam Style". We run YouTube searches for 8 different keywords: <Psy Gangnam Style Acoustic>, <Psy Gangnam Style Cover>, <Psy Gangnam Style Dance>, . . . , and <Psy Gangnam Style Remix>. Each search returns up to 1,000 results, which leaves us with a maximum of 8,000 UGCs for each music video. In the end, we collect 65,628 related UGCs for the top 100 music videos.

3.3 Video Attributes

Each video is associated with the following attributes: title, description, uploaded time, username of the uploader, location of the uploader, and the total view count. These attributes can be automatically extracted using the YouTube API. We also tagged the genre of the music for the top-100 most viewed videos. Music genre was automatically retrieved from the Wikipedia page for each song, where one song may belong to multiple genres.

4. CHARACTERIZING UGC UPLOADING BEHAVIOR ON YOUTUBE

We study three aspects of UGC uploading behavior on YouTube: what kind of UGCs related to popular music videos are uploaded, where, and when.

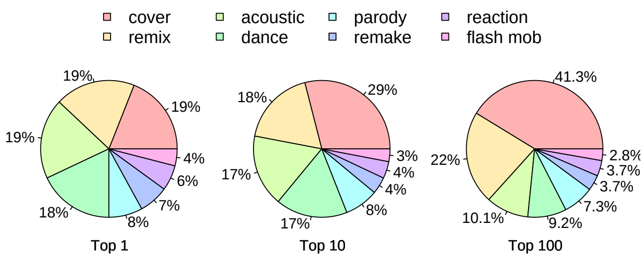


Figure 3: Frequency Distribution by UGC Type

4.1 What

First, we examine the types of UGCs and their distribution depending on the genre of the original music videos. Note that these UGCs that we analyzed are related to the top-most viewed music videos. Figure 3 shows the frequency distribution of UGC types, for top 1, top 10, and top 100 music videos. Overall, cover, remix, acoustic, and dances are prevalent, while reaction and flash mob types occupy only a small portion of the sample. Figure 2 shows the distribution of UGC types depending on the genre of the music it is dealing with. We find that irrespective of the genre, cover is the most popular type of UGCs. Remix UGCs are relatively prevalent for dance and electro-pop music, but not as popular for other music genre. As expected, R&B music videos spawned the least amount of the dance UGCs and the rap music videos are associated with only a few acoustic UGCs. The analysis results show that the UGC creation and uploading behavior varies across different genres, which indicates that UGC type analysis can be considered as a possible feature in identifying the genre of the original music video. In the following subsections, we analyze the temporal and spatial properties associated with UGCs across different types and genres.

4.2 Where

We now study the geographical diversity of UGC creating activity on YouTube. We start by describing the locations of top 100 most viewed original videos. We regard the location of a video uploader as the location of the music video, which is given in the standard two-letter ISO 3166 country code format. The top 100 music videos are generated from 6 different countries, United States (87), Great Britain (8), Australia (1), Italy (1), Korea (1), and Spain (1), and 1 unknown location. We notice that the majority of them were generated from the United States (US). Now we study the location of the related UGCs. The UGCs related to the top-1 original video, "Gangnam Style", were generated from 111 different countries. Top 10 and top 100 original videos were generated from 159 and 208 different countries, respectively. Figure 4 illustrates the countries that participated in creating the UGCs related to top-1, top-10 and top-100 music videos. Note that the set of countries that participated in UGC production for the top-1 video is a subset of that of the top-10 videos. The total number of countries registered as a part of ISO 3166 Country Codes is 249, which indicates that the UGCs related to top 100 most viewed music videos are uploaded worldwide. In other words, despite the

$LOC_{original}$	LOC_{ugc}	Frequency
United States	United States	15,588
United States	United Kingdom	4,983
United States	Canada	2,469
United Kingdom	United States	2,242
United States	Germany	2,081

Table 1: Top 5 Frequent ($LOC_{original}$, LOC_{ugc}) Pairs

fact that the top 100 popular music videos on YouTube were created in only 6 different countries, they triggered interests from worldwide, encouraging user participation from almost every part of the world. Yet we find that the US is the most active in creating UGCs, regardless of the UGC types. 29.2% of the UGCs we collected were made in the US.

Furthermore, we explore how the location of the original video ($LOC_{original}$) is associated with that of its UGCs (LOC_{ugc}). 680 unique pairs of $LOC_{original}$ and LOC_{ugc} were found. Table 1 shows the top 5 most frequent pairs. The largest number of UGCs (15,588 UGCs), which are uploaded from the US, originated from the US itself. Also, 4,983 UGCs uploaded from the UK are derived from the US music videos. For the top 5 pairs, western countries are usually associated with each other.

We now examine the music videos on the top-100 list individually, and compute the number of distinct countries that participated in creating the UGCs based on each video. For example Psy's "Gangnam Style", the top most popular music video on YouTube, caused participation from 111 different countries. The second most popular music video, Justin

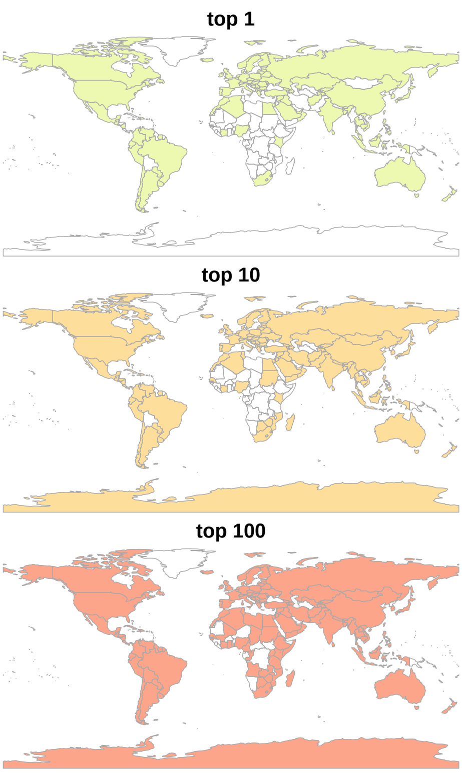


Figure 4: Countries Participated in UGC Uploads

$LOC_{original}$ (MV Count)	Average No. of LOC_{ugc}	Total No. of LOC_{ugc}
United States (87)	51	194
United Kingdom (8)	37	123
South Korea (1)	110	110
Australia (1)	98	98
Spain (1)	82	82
Italy (1)	73	73
Total (99)	54	208

Table 2: Number of $LOC_{original}$

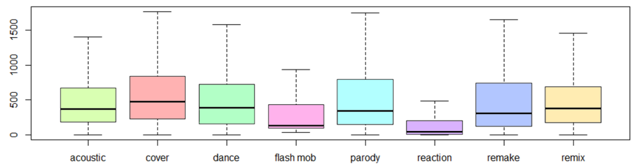
Bieber's "Baby", led to creation of UGCs from 92 different countries. We group the videos based on $LOC_{original}$ and compute the average number of distinct LOC_{ugc}s. The average numbers are shown in Table 2 in the second column. On the third column, we aggregate all the videos based on $LOC_{original}$ and find the sum of all distinct LOC_{ugc} for each $LOC_{original}$. As mentioned earlier, US and UK have more than one video on the top-10 list while the other countries appear only once on the top list. Thereby, the average and total number of LOC_{ugc} are the same for the last 4 rows of the table.

In total, the top-100 music videos yielded participation from 54 different countries on average (min=1, max=110, sd=30), confirming the idea that popular music videos and UGCs are globally associated. We observe the biggest diversity for South Korea, the home of the top-1 music video. Even other countries generated participation from a considerable number of countries, indicating that consumption of popular music on YouTube is an international activity. We further question whether a predominantly large number of distinct countries participating in the UGC production for South Korea is because "Gangnam Style" is ranked the top on YouTube or if this is just coincidental. We test if a correlation exists between the popularity of a music video, measured by its view count, and the video's "global" fame, measured by the number of LOC_{ugc}s. We find that the relationship between the two is statistically significant ($p<.01$) but its effect is weak (Spearman's ρ=0.27).

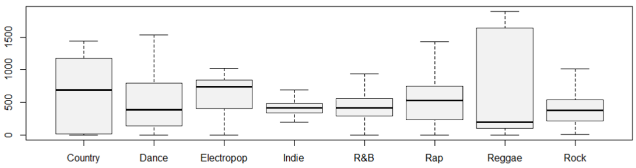
4.3 When

In this section, we look into when the UGCs are uploaded after their original videos were released. Let us refer to the date when an original music video was uploaded on YouTube as $DATE_{original}$, and the date when a related UGC was uploaded as $DATE_{ugc}$. For each UGC, we compute the time difference ($DATE_{ugc}$-$DATE_{original}$), which tells us how long it took for the UGC to be uploaded after the release of the original music video. We average the measures for each music video on the top-100 list, and find that it takes more than a year on average (491 days).

For a more detailed analysis, we investigate whether there is a difference in the UGC upload time depending on the types of UGC (Figure 5(a)) and on the genres of the original music videos (Figure 5(b)). Figure 5(a) shows that the reaction UGCs take the least amount of time. Reaction videos are the recent trend on YouTube, where the users film themselves watching a music video and express their feelings and comments while watching it. Since most reaction videos are intended to capture the first impressions, they can be cre-



(a) Elapsed Days by UGC Types



(b) Elapsed Days by Original Music Genre

Figure 5: Elapsed Time between Original MV Upload and Related UGCs Upload (in Days)

ated immediately after the original video is uploaded and do not require any preparation in advance. However, the other types of UGCs require some time for the users to familiarize themselves with the original music videos such as learning the dance moves or rearranging the songs and instrumentals. Figure 5(b) shows the reggae music UGCs take the least amount of time but have high variability. Also, the UGCs of dance music videos are uploaded relatively quickly.

We further study when the UGCs are spread out to other countries to understand when and where the UGCs are uploaded. Table 3 shows the top 5 quickest uploaded non-local UGCs. The fastest upload of UGC occurred within a week from the day the original was uploaded. This UGC was uploaded from Isle of Man after the original video from the United Kingdom was uploaded. We posit that this is due to the cultural and spatial proximity between the two countries, although further study is necessary to validate the hypothesis. Table 4 shows the locations of the top 100 original music videos and the average number of days (AVG($DATE_{ugc}$-$DATE_{original}$)) it took for the UGCs to be uploaded. We

$LOC_{original}$	LOC_{ugc}	Elapsed Days
United Kingdom	Isle of Man	7
United States	Zambia	9
United States	Chad	17
South Korea	Cayman Islands	20
South Korea	Liechtenstein	30

Table 3: Top 5 Quickest ($LOC_{original}$, LOC_{ugc}) Pairs

$LOC_{original}$	Avg. Elapsed Days
Spain	741
United States	524
Australia	485
United Kingdom	451
Italy	114
South Korea	111

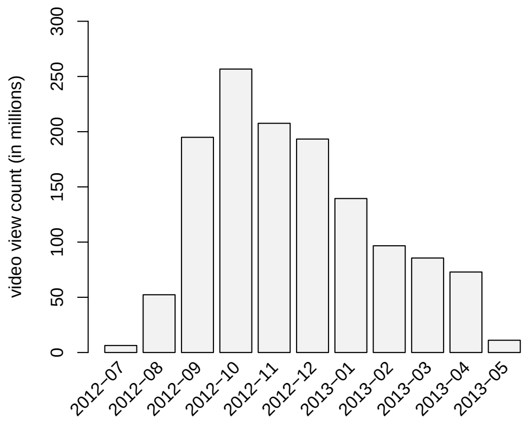
Table 4: Average Time between Original MV Upload and Related UGCs Upload (in Days)

find that it takes the longest on average for the US music videos. It may be due to the uneven data distribution since the US has the most original videos and UGCs. A notable point is the relatively small time difference for the South Korean music video, "Gangnam Style". It shows that the UGCs on "Gangnam Style" was created in more than 100 countries within four months after the official "Gangnam Style" music video was uploaded.

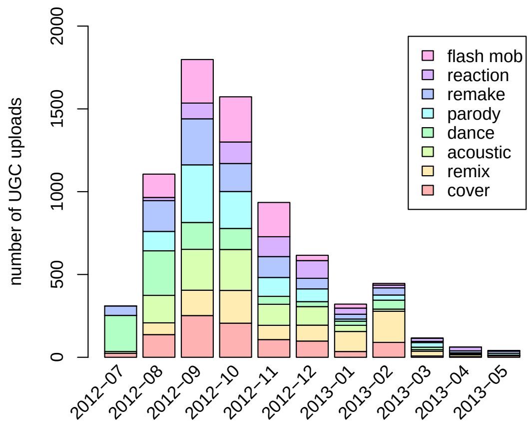
5. YOUTUBE VIDEO POPULARITY AND UGC UPLOADING BEHAVIOR

In this section, we explore the relationship between the popularity of an official music video with respect to the volume of related UGCs being uploaded on YouTube. As explained earlier, "related" UGCs refer to the recreated content of the original music video including acoustics, covers, dances, flash mobs, parodies, reactions, remakes, and remixes.

We examine the top most viewed music video on YouTube, "Gangnam Style" by a Korean pop-star Psy. In particular, we question whether there is a correlation between the



(a) View Counts for "Gangname Style" MV over Time



(b) Number of Related UGC Uploades over Time

Figure 6: Original Music Video View Counts vs. UGC Upload Frequency

view count of Gangnam Style music video and the number of UGCs related to Gangnam Style. Figure 7 shows the change in the number of view counts of the official "Gangnam Style" music video (a), and the change in the number of "Gangnam Style" UGCs being created (b) over time. The original "Gangnam Style" music video has reached its maximum number of views in October 2012 with a sudden jump in the preceding month and a gradual decrease afterwards. We can immediately notice from the graphs that the changes in the number of related UGCs uploaded closely resemble the growth pattern of the "Gangnam Style" music video view count; a sudden leap in the number of uploads lead to a peak in September 2012, followed by a gradual decrease afterwards.

Moreover, we observe that the change in the number of UGC uploads precedes the change in the music video view counts, which suggests the potential of utilizing the information from re-created UGCs for predicting the popularity of the original content. We apply the econometric technique of bivariate Granger causality analysis [6] to the monthly time series of UGC upload volume and original MV view count. Through this analysis, we can test whether the values of time series X will exhibit a statistically significant correlation with the future values of time series Y. In other words, X is said to Granger-cause Y if Y can be predicted better by using the histories of both X and Y than using the history of Y alone. The analysis results in a p-value below 0.001, with a lag value of 1 month. It allows us to reject the null hypothesis that UGC upload time series does not predict the popularity (view counts) of the original music video, indicating that UGC upload volume Granger-causes the original video view count.

6. CONCLUSIONS AND FUTURE WORK

In this study we explored the phenomenon in which YouTube users create and upload videos that are the "re-creations" of the popular music videos uploaded by the official record labels or agencies. We identified different types of user-generated content (UGC) including covers, dances, parodies, reactions, flash mobs, etc., and found that the UGC uploading behavior is different across different types of UGCs as well as the genres of the original music videos. In addition, we found that sharing of UGCs is a global activity, drawing participation from a world-wide audience, and that it is a long-lasting activity, where it takes more than a year on average for UGCs to be uploaded.

The second phase of our analysis involved correlation analysis between the volume of related UGCs and popularity of the original content. We found that the extent to which UGCs are uploaded on YouTube is positively correlated with the popularity of the original video, where popularity is measured by the number of view counts. It implies that the analysis of UGC uploading behavior can help one in understanding the popularity of the original content. We supplemented this hypothesis through a Granger-causality test, where we provided evidence that the changes in the volume of related UGCs uploaded on YouTube is predictive of the changes in the popularity of the original music video. As also discussed in Cha et al's work [4], predicting the future popularity of content is important for service providers in terms of devising effective proxy caching strategies to accelerate service requests, which could ultimately lead to overall satisfaction of users of the service.

Due to the YouTube policy, it was not possible to crawl all the related UGCs of the original music videos. Thereby our analysis is based on the popular music videos and the relevant UGCs with the highest view counts. Although this could be a possible source of bias in the analysis results, we believe it is minimal because these top popular videos account for majority of the traffic on YouTube [4]. For future work, we plan to collect user ratings and comments along with the view counts as additional indicators of video popularity, and study the relationship between them and the growth of UGC uploads.

7. ACKNOWLEDGMENTS

This work was supported in part by WCU (World Class University) program under the National Research Foundation of Korea and funded by the Ministry of Eduation, Science and Technology of Korea (No.R31-30007), and in part by the National Research Foundation of Korea grant funded by the Korean government (MSIP) (No. NRF-2009-0081365).

8. REFERENCES

[1] Y. M. Baek and I. Shin. Cultural distance and intercultural consumption of korean pop music on social media. In *annual meeting of the International Communication Association*, London, England, 2013.

[2] Y. Borghol, S. Mitra, S. Ardon, N. Carlsson, D. Eager, and A. Mahanti. Characterizing and modelling popularity of user-generated videos. *Performance Evaluation*, 68(11):1037–1055, Nov. 2011.

[3] A. Brodersen, S. Scellato, and M. Wattenhofer. Youtube around the world: Geographic popularity of videos. In *Proc. WWW*, pages 241–250, New York, NY, USA, 2012. ACM.

[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 *Proc. IMC*, pages 1–14, New York, NY, USA, 2007. ACM.

[5] F. Figueiredo, F. Benevenuto, and J. M. Almeida. The tube over time: Characterizing popularity growth of youtube videos. In *Proc. WSDM*, pages 745–754, New York, NY, USA, 2011. ACM.

[6] C. W. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438, 1969.

[7] H. Pinto, J. M. Almeida, and M. A. Gonçalves. Using early view patterns to predict the popularity of youtube videos. In *Proc. WSDM*, pages 365–374, New York, NY, USA, 2013. ACM.

[8] S. Siersdorfer, S. Chelaru, W. Nejdl, and J. San Pedro. How useful are your comments?: Analyzing and predicting youtube comments and comment ratings. In *Proc. WWW*, pages 891–900, New York, NY, USA, 2010. ACM.

[9] A. Susarla, J.-H. Oh, and Y. Tan. Social networks and the diffusion of user-generated content: Evidence from youtube. *Information Systems Research*, 23(1):23–41, Mar. 2012.

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