On the Geographic Patterns of a Large-scale Mobile Video-on-Demand System

Zhenyu Li*, Gaogang Xie*, Jiali Lin*, Yun Jin†, Mohamed-Ali Kaafar‡, Kavé Salamatian†
*ICT, CAS, China, †PPTV, China, ‡ INRIA France, National ICT Australia § Université de Savoie, France
{zyli,linjiali, xie}@ict.ac.cn, yunjin@pptv.com, mohamed-ali.kaafar@inria.fr, kave.salamatian@univ-savoie.fr

***Abstract*—The widespread availability of smart mobile terminals along with the ever increasing bandwidth capabilities has promoted the popularity of mobile Internet video systems. Understanding the geographic features of mobile content consumption is of an extreme importance for the design and the performance optimization of a mobile video delivery system. This paper is a first step towards characterization of the geographic patterns of a large-scale commercial mobile video-on-demand (VoD) system, by measuring both uniformity and intensity of geographic interests on videos. In particular, we identify a geographical concentration effect of views for individual videos, which is however dependent on video popularity. We also analyze the temporal evolution trends of the geographic popularity which reveal distinct behavior of popular and non-popular videos. While the set of locations that contribute to most of the views of non-popular videos largely varies, the daily geographic popularity distribution of popular videos closely follows the distribution of global traffic and remains stable. We also examine the impact of content type and viewing sources on the geographic features of mobile videos consumption, and the correlation between content similarity and geographic locality. Finally, we provide insights into the implications of our findings.**

I. Introduction

The past decade has witnessed the widespread use of mobile and wireless networks. With high available bandwidth and an increasing popularity of mobile terminals, watching videos online from mobile terminals becomes ubiquitous. Today, content providers compete to offer a large selection of video content which would attract the largest fraction of mobile users and hence increase their audience and revenue. This leads to an ever-growing popularity of mobile video services and to an increasing traffic volume [1], which in turn creates a thrust among the research community and industry to investigate issues related to the adaptivity in mobile content delivery and distribution.

In such a context, understanding peculiar features of mobile video geographic popularity is particularly important. The extent to which the popularity of commercial mobile videos might be dependent on user geographic locations, if learned correctly, can lead to optimized content placement strategies, and in turn, to improved service and reduced costs [6].

Recent works suggest that geolocation affects user-generated content (UGC) consumption and content popularity [12] [4]. However, to date the geographic patterns of commercial mobile video systems remain not well understood. This paper seeks to address this gap. Our work differs from previous works on UGC content [12] [4] from at least three perspectives. First, we study a commercial mobile video system which as opposed to UGC content (such as videos uploaded on YouTube), only comprises commercial video content that targets a global audience. Second, while video discovery in UGC heavily relies on online social networks as shown in [4], we observe in the studied mobile VoD system that users rely on the main app to discover and watch videos. Finally, while previous studies do not distinguish between mobile and regular Internet users, we focus here on the behavior of mobile users towards video on demand consumption. These lead to distinct geographic popularity patterns as shown in this paper.

Our study is based on a unique dataset collected from PPTV, a leading video provider in China. The dataset consists of about 87 million viewing requests generated by more than 3.5 million mobile users from not only mainland China but also overseas regions. By measuring both uniformity and intensity of geographic-based interests of individual videos, we examine several geographic features of video consumption, including the impact of content type and viewing source, and the correlation between content similarity and the geographic locality of video popularity. We then study the temporal evolution of the geographic-based popularity for different categories of content. In essence, this paper provides the following findings.

- We find that about 40% of videos have more than 70% of their views in 3 (of 33) provincial locations, showing an effect of geographical concentration. Such a concentration effect however is dependent on video popularity. Besides, compared to the geographic patterns of UGC content observed through YouTube videos in [4], VoD content shows less local geographic popularity due to the nature of content.
- We find that the viewing sources from which users discover videos have a substantial impact on the video geographic popularity. In particular, videos accessed through the search portals are likely to be local from a geographic popularity perspective. Other sources however yield a global, *i.e.* unlocalized, geographic video popularity. One of our salient findings is that similar content exhibits similar popularity distributions across geolocations, especially for TV series videos.
- Depending on the level of video popularity, we observe distinctive geographic popularity trends over time. Non-

978-1-4799-3360-0/14/$31.00 ©2014 IEEE

978-14799-3360-0/14/$31.00 ©2014 IEEE

popular videos have a stable daily popularity and attract most of their views from a set of a few locations, the diversity of which however varies greatly from day to day. Popular videos on the other hand experience a peak daily popularity shortly after their upload and their daily geographic popularity distributions closely follow the distribution of global traffic. Such temporal trends are greatly different from that of UGC content illustrated in [4].

- We observe that although content categories have a limited impact on the overall geographic popularity of videos, the evolution trend of geographic popularity over time significantly varies according to the video category. For example, regardless of the video popularity, videos belonging to the news category have a global geographic popularity within the first few days and then suddenly exhibit a view concentration in only a few locations.
- We finally present insights on possible applications of our findings for content placement, online advertisement and mobile applications design. We envision for example that popular mobile videos can be replicated on the CDN servers following the geographic distribution of the global traffic, an information that can be easily accessed by the CDN provider. Non-popular videos however require the usage of other features to predict the locations where users have specific interests in them.

The remainder of this paper is organized as follows: Section II describes the dataset and the methodology in use to identify video geographic locations. In Section III, we perform an in-depth analysis of geographic popularity. Section IV examines the temporal evolution of the geographic popularity. We discuss the implications of our findings in Section V. Section VI surveys the related work and we conclude in Section VII.

II. Dataset and Methodology

A. Dataset

Our dataset is collected from the PPTV mobile VoD servers' logs, during the two first weeks in December 2011. The logs record the viewing requests originating from mobile terminals (*e.g.* iPhone, iPad, Android Phones, etc.). We observed that more than 98% of the views from mobiles were generated by the PPTV mobile devices dedicated software (*i.e.* mobile app, as opposed to web browsers). We then only consider the viewing records from the PPTV software [2].

For each view, a log entry contains a timestamp indicating the time when the request was issued, an identifier of the requested video, the client unique identifier (generated when the PPTV app is installed on the mobile device), the client provincial geolocation, the category of the video (amongst 22 categories, *e.g.* movies, TV series, cartoon, variety shows), the view duration and the referring source illustrating how the user reached the video (*e.g.* video search, category page, user subscription). There are 33 geolocations in total, including 31 provinces in mainland China, Hong Kong Macau and Taiwan (referred as HMT in the following), and other world-wide countries (referred as Oversea). Overall, the dataset contains

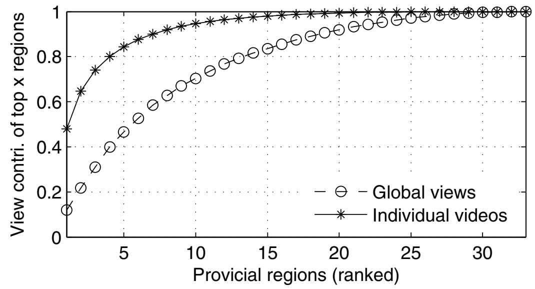


Fig. 1: Geographic popularity distribution for individual videos and global traffic

86,521,403 log entries, with a proportion of 92.58% access via a WiFi connection and 7.42% via a 3G connection, and with a proportion of 42.08% access from tablets (*i.e.* iPad and Android pad), 57.91% from smart phones (*i.e.* iPhone and Android phone). These views were generated by 3,759,129 users watching 427,316 unique videos.

B. Measuring Geographic Popularity of Video

Let v_i denote video i's aggregate number of views (*i.e.* popularity) during the observation period and v_{il} its popularity at a particular location l. We are interested in how uniform the requests of a video are distributed across geographic locations (*i.e. uniformity* of interest) and to what extent the requests come from a few top locations (*i.e. intensity* of interest). Inspired by [4] [5], we use the view *entropy* of a video to measure the geographic uniformity and the *view focus* to measure the intensity. Specifically, the entropy of a video i is defined as follows:

$$E(i) = \frac{-\sum_{l=1}^{N} \frac{v_{il}}{v_i} \times log \frac{v_{il}}{v_i}}{log(N)} \tag{1}$$

where N is the number of geolocations, *i.e.* $N = 33$. The entropy is a number between 0 and 1. The smaller the entropy value is, the more biased the popularity distribution towards a few locations is, while a larger value indicates a more uniform distribution. The view focus of a video i is defined over the top k locations that attract most of the views of the video as follows:

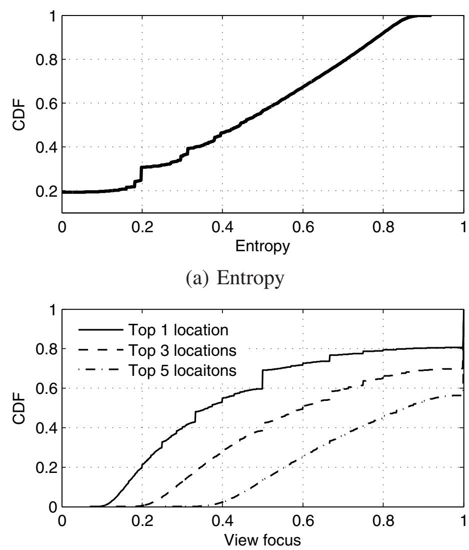
$$VF(k) = \frac{1}{v_i} \sum_{l=1}^{k} v_{il} \tag{2}$$

These two metrics can be easily extended to a particular day t by replacing v_i with v_i^t and v_{il} to v_{il}^t, where v_i^t is the number of views of video i during day t.

III. Geographic Popularity of Videos

A. Overview of video geographic popularity

We first examine whether there is a concentration effect on the geographic popularity by comparing the geographic popularity distribution of individual videos and the distribution of global traffic. Specifically, for a given video, we rank the locations according to the video's popularity in each



(a) Entropy

(b) View Focus

Fig. 2: Cumulative distribution of entropy and view focus

province and calculate the cumulative distribution of views per individual video over the sorted locations. Similarly, for the global traffic, the locations are ranked in a descending order according to the number of views of *all videos* in the dataset. Figure 1 plots the average of cumulative distribution values over individual videos in function of the top locations. We also show the distribution for the global traffic in function of the contributed locations.

We observe an unbalanced distribution of global traffic: the top 1 location generates 13% of the views and the top 10 locations generate more than 70% of the views. The disparity in mobile population and wireless infrastructure among locations explains such an unbalanced distribution. The distribution for individual videos is even more skewed, as on average the top 1 location generates up to 50% of a video's views and only a few locations already contribute the vast majority of the views (*e.g.* top-4 locations generate up to 80% of a video's views). This observation clearly indicates a high concentration effect of geographic popularity of individual videos.

B. Quantitative measure of geographic popularity

Figure 2 depicts the cumulative distribution (CDF) for view entropy and view focus over videos. In Figure 2a, we observe that 55% of the videos experience an entropy less than 0.5, which suggests a biased geographic popularity distribution towards a small set of locations. Notably, 20% of the videos have a view entropy of 0, which indicates that these videos attracted views from a single location.

Figure 2b depicts the CDF of the view focus. The figure shows that for 40% of the videos, half of their views come from a single location, 70% of their views from 3 locations, and all their views only from 5 locations. The CDF is enlightening and confirms the geographic concentration of the mobile videos' views.

Compared with the geographic popularity distribution of YouTube videos shown in [4], where 40% of the videos attract 80% of their views from a single location, our results here show a smaller bias towards geographic distributions. One possible explanation of such a difference is the nature of the content itself, which in the case of YouTube is user-generated content and as such different content is distributed in various languages rather intended at upload time to be specific to a user population. On the other hand, in a VoD system, content is provided by commercial providers in order to reach the largest audience of viewers. Our results however still demonstrate the high concentration of interests of mobile users in specific content, while showing the ubiquitous nature of mobile VoD.

In the following analysis, we focus on the view focus of the top-3 locations, as the top-3 locations generate more than 60% of views for as many as 50% of videos.

C. Impact of popularity and content category

TABLE I: Statistics on video category

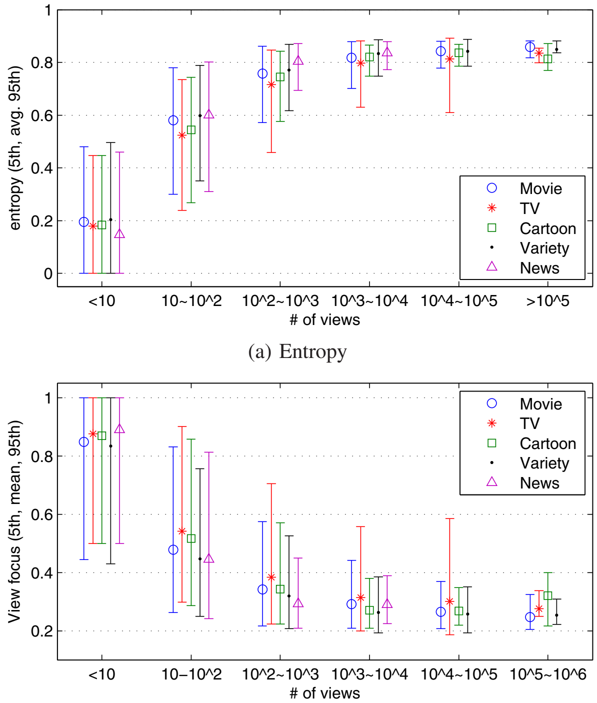
category	videos(%)	views(%)
Movie	6.3	23.0
TV series	37.7	40.2
Cartoon	19.6	21.9
Variety shows	14.0	10.7
News	5.1	2.7

Videos with different popularity or of different content categories might experience different geographic popularity distributions. PPTV offers videos of 22 content categories. However, as shown in Table I, most of the videos/views come from the top 5 categories. We bin videos of each of these 5 categories according to the popularity (*i.e.* number of views). Figure 3a and Figure 3b plot the 5th percentile, mean and 95th percentile of the entropy and view focus (top 3 locations) for each bin, respectively.

In general, the entropy increases with the growth of popularity, while the view focus decreases, which suggests a more balanced geographic distribution for popular videos. However, there seems to be a limit for both entropy and view focus as the mean values show steady variations when the video popularity reaches 1,000. Comparing the entropy and view focus among categories, we find that the mean values of different categories exhibit a small variation. This shows that on average, videos of different categories do not exhibit a distinct geographic popularity distribution. However, we observe that some very popular videos of TV series show a highly skewed distribution. For example, about 5% of the videos from TV series category with a popularity higher than 10,000, have 60% of their views from only 3 locations. We checked these videos and found that they are in Cantonese language and produced in Hong Kong, and as such most of the requests were likely generated by users from a limited number of Cantonese-spoken locations.

D. Impact of viewing source

To enable improved user experience, Internet video providers develop various ways for users to discover videos. In the following, we focus on the impact of viewing sources



(a) Entropy

(b) View Focus

Fig. 3: Impact of popularity and content type

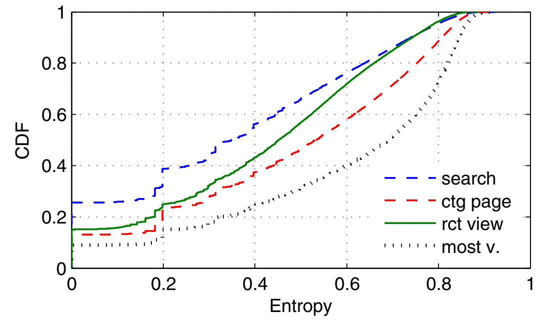


Fig. 5: Impact of viewing source on entropy

on the geographic popularity. The PPTV mobile app provides 8 different discovery methods: the category page (*i.e.* the front page of each category), search portal (of the mobile app), recent views (of the user), recommendation to user, favorites of user, most viewed videos (in the system), homepage of the app and subscription of user.

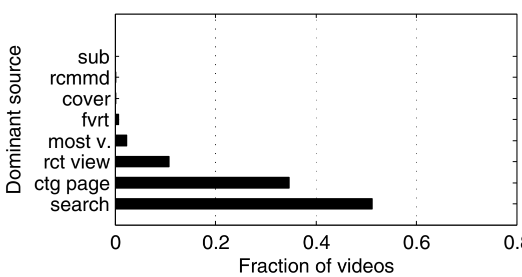


Fig. 4: Views from each source

We define the *dominant viewing source* of a video as the viewing source from which most of the views of the video originate. Figure 4 shows the distribution of the dominant sources for individual videos. The search portal is the dominant source for more than half of the videos, which is different from YouTube where the most popular dominant source is the Related Video [15] (*i.e.* recommendation-based videos) .

In Figure 5 we study the impact of the top 4 dominant view sources on the geographic popularity, as these 4 sources represent more than 98% of the videos' sources. We observe that the view source has a significant impact on the geographic popularity. It is interesting to note that videos with viewing sources being dominated by the search portal have a more local geographic popularity. In essence, users discover most of their specific (*i.e.* non uniform across all locations) interests by explicitly searching for them. We note that the Category Page on the other hand contributes to a more global geographic popularity.

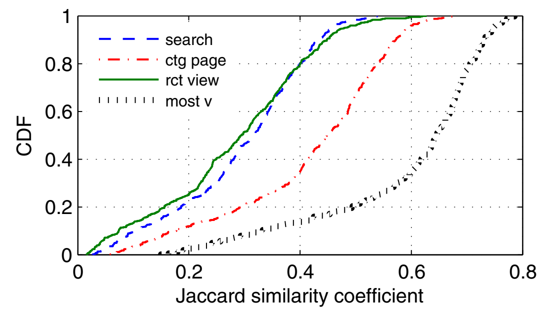


Fig. 6: Jaccard interest similarity between any two locations

To confirm that Search is more often used by users when discovering videos of their particular interests, we divide videos into four groups based on the dominant viewing source. For a particular group of videos, we associate each geolocation l_i a set $S(i)$, containing the videos of this group that users from l_i requested. We then compute the Jaccard similarity coefficient of the two locations l_1 and l_2 as follows:

$$J(l_1, l_2) = \frac{S(l_1) \cap S(l_2)}{S(l_1) \cup S(l_2)} \tag{3}$$

The Jaccard coefficient lies in between 0 and 1 and a value greater the 0.5 is typically considered to illustrate a considerable similarity. Figure 6 depicts the cumulative distribution of the obtained similarity coefficients. Considering the videos dominated by the search portal, a very limited number of location pairs have a similarity coefficient greater than 0.5, which shows that the search portal is used by users of different locations for discovery of videos of their specific interests. The coefficients of the Category Page and Most Viewed are much higher, indicating as expected that videos dominated by these two viewing sources are more globally popular.

E. Content and geographic locality

As shown previously, videos might attract views from a few locations only (local) or from a global audience. An interesting

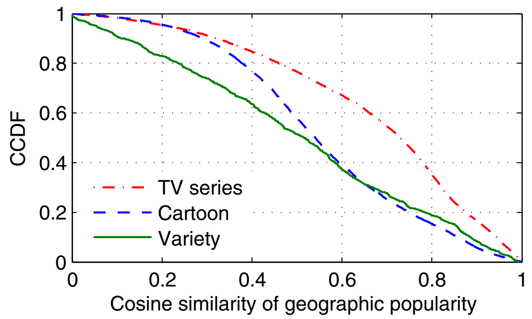


Fig. 7: Location similarity of videos with similar content

question is whether videos of similar content share similar popularity distribution across geographic locations [6]. To this end, we associate with each video i a geographic popularity vector L_i, in which a typical element is $< l, v_{il} >$ (v_{ij} is the number of views of video i from the location l). The similarity of geographic popularity distribution of two videos i and j is then measured using the cosine similarity as follows:

$$sim(L_i, L_j) = \frac{L_i \cdot L_j}{|L_i||L_j|} \tag{4}$$

The similarity value varies from 0 and 1 where a larger value indicates a higher similarity. Again, a cosine similarity greater than 0.5 is typically considered to illustrate a significant similarity. In the PPTV mobile VoD system, videos of similar content, for instance, episodes of TV series, cartoons or variety shows, are marked with the same video group identifier. For any two videos with the same identifier, we compute the cosine similarity of their geographic popularity vectors and plot the CCDF (complementary cumulative distribution function) of the results for TV series, cartoon and variety shows, as videos with similar content from these 3 categories represent the vast majority (97%) of similar content.

We observe that independently of the video category, more than half of the video pairs experience a location similarity higher than 0.5. Interestingly, videos of TV series exhibit a higher location similarity than the videos of other two categories, with more than 30% of the video pairs of TV series showing a similarity greater than 0.8. These results implies that videos of similar content share similar popularity distribution across geolocations, which can be leveraged to design an efficient geographically distributed content delivery system as illustrated in [6].

F. Summary

Our main findings in this section are summarized as follows.

- Individual videos exhibit a more local geographic popularity towards a few locations, compared with the global traffic of all videos. In particular, there are as many as about 40% of videos attracting 70% of their views from 3 out of 33 provincial locations.
- Regardless of the content category, non-popular videos show a more biased geographic popularity distribution than the popular ones. However, some particular categories of very popular videos, *e.g.* TV series, also show a very skewed geographic distribution.
- Videos with viewing sources being dominated by the search portal generally lead to a local geographic popularity, while videos which referrals are dominated by "Category Page" and "Most Viewed" show a balanced popularity distribution across geolocations.
- Videos of similar content show very similar popularity distributions across geolocations.

IV. TEMPORAL EVOLUTION OF GEOGRAPHIC PATTERNS

This section examines the evolution of geographic popularity distribution for individual videos over time. Such an evolution reveals how the sets of locations that generate requests to individual videos vary over time, which is of great importance for video placement in distributed content delivery networks. As the exact upload time of videos is not available in our data set, we provide an estimate of the upload time in the following way: we consider a video as being a fresh upload if it was not viewed in the two preceding days (*i.e.* from Dec. 1 to Dec. 2, 2011) and mark its upload time to be the first time it was viewed in our dataset. The videos uploaded from Dec. 3 to Dec. 5 and viewed at least once every day since upload were selected for analysis. We only show here the results for the videos uploaded on Dec. 3 (Saturday) as videos uploaded on the two other days exhibit very similar trends.

TABLE II: Videos updated on Dec. 3

cat.	TV	cartoon	variety	news	others
%	47.7	19.7	13.3	9.9	9.4

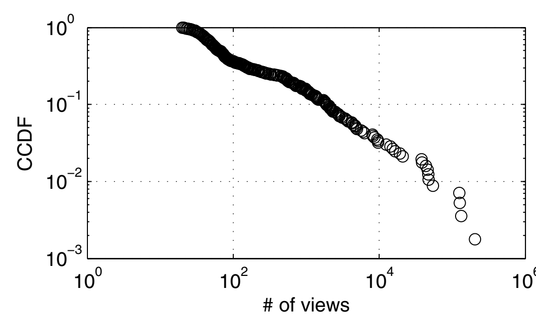
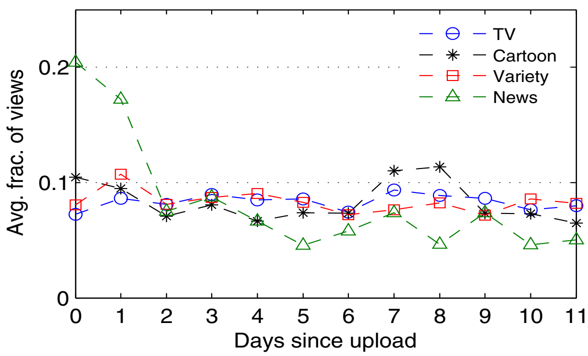


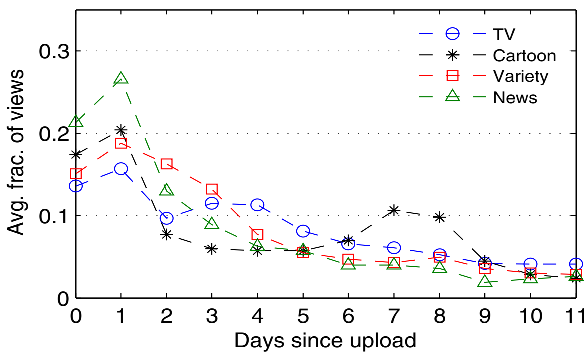
Fig. 8: Power-law popular distribution of videos uploaded on Dec. 3

In total, 566 videos are under consideration, which are summarized in Table II. Almost 70% of the videos belong to TV series and cartoon categories, which often contain dozens of episodes uploaded every day. Variety shows are scheduled for upload during weekends, and the news videos are video clips of hot and timely events.

Figure 8 shows the CCDF of the popularity for these videos in a log-log scale. Although the videos were uploaded on the same day, their popularity roughly follows a power-law distribution. While lots of videos only attract dozens of views,

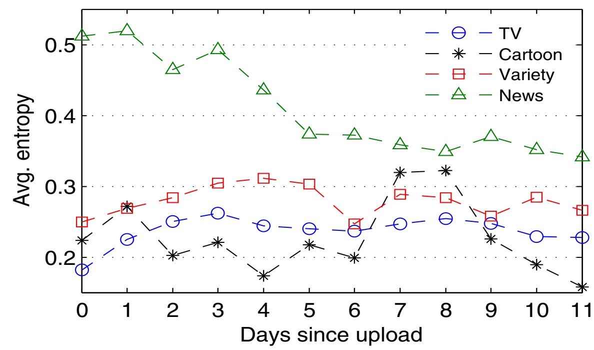


(a) non-popular videos

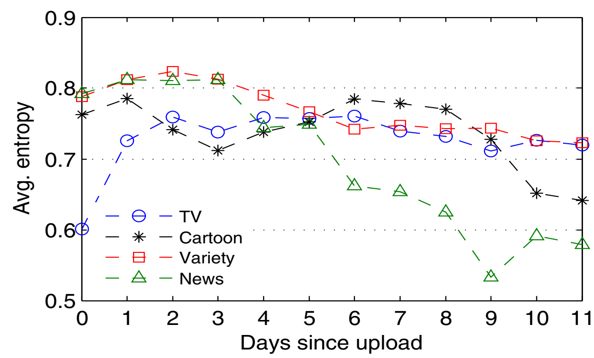


(b) popular videos

Fig. 9: Video popularity evolution trend



(a) non-popular videos



(b) popular videos

Fig. 10: Entropy evolution trend

a few get more than 100,000 views in 12 days since the upload. Given the large disparity of popularity of individual videos, we divide videos into two groups according to their popularity: *non-popularity videos* with less than 1,000 views and *popular videos* with more 1,000 views.

A. Popularity evolution

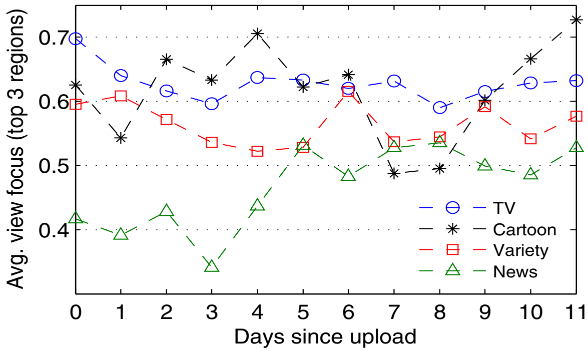
We first examine the video popularity evolution over time. For each video, we compute the normalized number of views for every day since its upload time. The average number of views over all videos of the same category is then computed and shown in Figure 9, where day 0 is the day when the video was uploaded.

We observe different popularity evolution trends depending on whether we consider popular or non-popular videos. Except the "news" videos category, the daily popularity of non-popular videos of the other 3 categories is very steady. However, all types of popular videos have a peak of popularity one day after their upload date. Subsequently, the popularity quickly drops over time to a low level. Comparing the temporal evolution trends among different categories, we make two notable observations. First, regardless of the video popularity, news videos exhibit a peak popularity within the first two days after the upload, and then experience a loose of interest much faster than other videos categories. This is to be expected and is due to the nature and timely interests into such kind of videos. Second, we observe an increase of interests in cartoon videos in the 7th and 8th days (weekend) since the upload. This is explained by the fact that the cartoon category videos are mostly watched by a young population of users during non-school days. While these findings are to be expected, it is worthwhile to notice that from a content provider perspective, these observations can be used to well-provision the servers while predicting the workload on a daily basis.

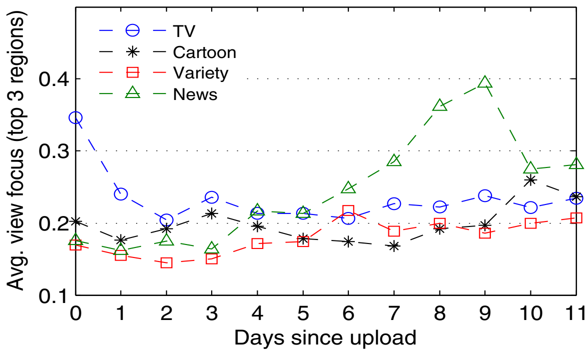
B. Geographic popularity evolution

Figure 10 illustrates the temporal evolution (in days) of the average entropy for different categories of videos. Non-popular videos (Figure 10a) have a much lower entropy over time, which suggests a concentration effect of their views in a few locations. Besides, the evolution of entropy follows the same trend of the temporal evolution of popularity shown in Figure 9: a higher popularity generally corresponds to a higher entropy. Interestingly, while we see a loose of interests in popular videos in the TV series category after the 1st day, the entropy still increases and then stabilized. In other words, such videos steadily attract views evenly from all locations.

Figure 11 depicts the view focus evolution trends. Here, we consider the view focus of the top-3 locations for videos. Note that we also compute the view focus of the top-1 location and similar trends were observed. The results show that the more locations a video attracts views from, the smaller the fraction of views from the top locations is. This finding is greatly different from the one of YouTube videos in [4], where the entropy and the view focus increases at the same time to their peaks, meaning that although users from more locations



(a) non-popular videos



(b) popular videos

Fig. 11: View focus (top 3 locations) evolution trend

request the videos, the requests from top locations increase even faster.

The difference can be explained by both the content nature and the impact of social diffusion. YouTube videos are produced by users with various background and they tend to attract most of their views from the locations where they are generated [4]. PPTV videos on the other hand are provided by content providers aiming to reach a global audience. Beside, YouTube videos greatly rely on social networks for content diffusion [4], which diffuses content in a small region first and then virally spread the content beyond the genesis location. This effect of social diffusion is also observed in Twitter for tweets geolocation popularity evolution [18]. Such an impact of social diffusion currently does not exist in the PPTV mobile VoD system.

Comparing the geographic popularity evolution trends among different categories, we observe that videos of the "news" category have a more global geographic popularity during the first few days and then quickly concentrate in a few locations only, independently of their level of popularity. Among the other 3 categories, videos of TV series show a more local geographic popularity. In essence, views of the news category show more timely interests than other type of content, and TV series provided on the VoD mobile platform are produced (and tailored) according to specific languages and users from a small set of locations. The sharp decrease of view focus for the non-popular cartoon videos in the 7th and 8th days is explained by the increase of interests on such videos from many locations in the weekend, as depicted in 9a.

C. Location stability analysis

We have observed that non-popular videos steadily attract most of the views from a limited number of locations, while popular videos tend to have a balanced view distribution across geolocations. However, from a video placement and caching strategy perspective, we are also interested in answering two follow-up questions. First, do non-popular videos consistently attract views from the same set of locations? Second, does the popularity distribution of popular videos across geolocations follow the global traffic distribution?

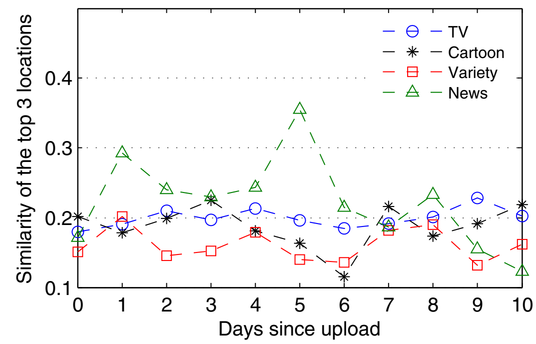


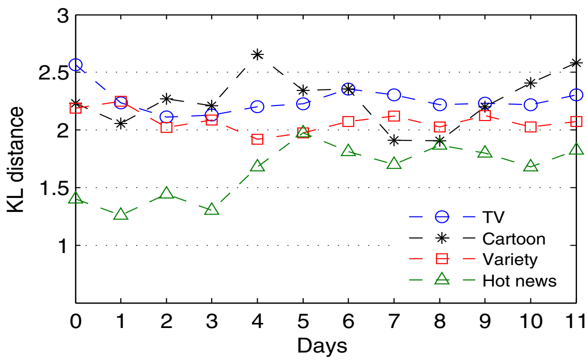
Fig. 12: Similarity of the top 3 locations for non-popular videos

To answer our first question, we measure for individual non-popular videos the overlap of the sets of the top 3 locations in two consecutive days using Jaccard similarity coefficient. We compute for each of the non-popular videos uploaded on Dec. 3 the daily location similarity and then compute the average similarity values over the videos of the same category. The results are shown in Figure 12 which depicts the average similarity values of the top-3 locations as a function of the days during the observation period. Surprisingly, we observe a low daily location similarity around 0.2 independently of the video type. This low similarity indicates that on average there is less than 1 location overlap in two consecutive days for such videos. Thus, although the daily popularity of non-popular videos remains stable, the sets of top locations generating most of their views experience a large variation.

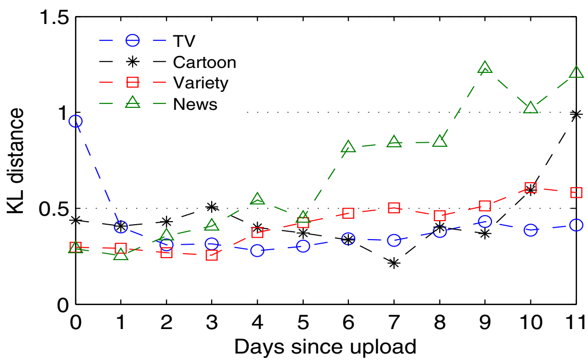
We answer the second question by measuring the divergence between the geographic distribution of global traffic and video daily geographic popularity distribution. To this end, we use the Kullback-Leibler (KL) divergence (distance) [7] to measure the distance between the two distributions, which is defined on two distributions P and Q as:

$$D_{KL}(P||Q) = \sum_{x \in X} P(x) log \frac{P(x)}{Q(x)} \quad (5)$$

The KL distance is a positive value only defined if $Q(i) = 0$ implies $P(i) = 0$. It represents the number of extra bits needed to code samples from P when using a code based on Q, rather than directly based on P. The smaller the value is, the closer



(a) non-popular videos



(b) popular videos

Fig. 13: Kullback-Leibler (KL) distance between daily geographic popularity distributions of individual videos and the distribution of global traffic

the two distributions are. In our context, X refers to the set of geolocations, P refers to the geographic popularity distribution of a video on a particular day and Q to the distribution of global traffic.

Figure 13b depicts the average KL distance values over videos of different categories for popular videos. Excluding videos of the "news" category, other types of videos have a KL distance between 0.3 to 0.5, which conveys that the daily geographic popularity distributions of individual videos closely follow the geographic distribution of global traffic. For comparison, we also show the KL distance values for non-popular videos in Figure 13a, which is almost always greater than 2, implying a different distribution than the global one. The reason is that non-popular videos attract views mostly from a few locations where users have particular interests on the content.

The distributions of individual videos following the distribution of global traffic is an important finding from the content placement perspective, since content providers could map the content on servers according to the global distribution, an easier to collect and more stable piece of information.

D. Summary

Our main findings in this section are summarized as follows.

- Non-popular videos tend to have a stable daily popularity and attract most of their views from a small set of locations. However, the set of locations generating most of the views vary greatly from day to day.
- Popular videos on the other hand experience a peak daily popularity shortly since their upload, and a fast subsequent popularity decrease. However, these videos attract views more uniformly from all locations even when the popularity greatly decreases. Interestingly, the daily geographic popularity distributions of individual videos closely follow the distribution of global traffic.
- Videos belonging to the News category obey to very different temporal evolution trends of geographic popularity, compared to the other categories of videos.

V. IMPLICATIONS

Although our findings are based on the PPTV mobile VoD system, the massive number of users and the large geographic coverage not only in mainland China but also in oversea countries allow us to gain insights on geographic popularity patterns for other similar online content platforms. Moreover, we figure out several implications of the findings in video placement, advertisement and mobile Internet video app design.

In a geographically distributed storage system, such as a hierarchical CDN network [16], a key problem is where to replicate videos. Our daily-based evolution analysis reveals that popular videos follow the similar geographic popularity distribution with the global traffic. Thus, they can be replicated on the top popular geolocations according to the distribution of global traffic, which can be easily obtained. However, it is much harder for non-popular videos to find proper geolocations to store the replicas, although they attract views only from a few locations every day, as the set of top locations change greatly from day to day. The popularity of videos can be predicted accurately within a few hours since upload as analyzed in [9].

We can also leverage the fact that videos with similar content exhibit a high similarity in geographic popularity to estimate the locations where a new video will be popular, and then store the videos on the servers in these locations as in [6]. One should also particularly notice the disparity among different video categories. For example, videos belonging to the "news" category exhibit different geographic popularity evolution trends compared to other categories.

Internet video systems greatly rely on the short advertisement before the start of playback for revenue. Our geographic popularity analysis results can be used for advertisement strategies. For example, advertisements that aim at a global audience can be correlated to the videos with a global popularity every day, such as popular variety shows or cartoon videos. The location-specified advertisements on the other hand can be played along with videos that are specially viewed by uses in particular locations. One can also take advantage of the increased interests on cartoon videos in weekends for purposely advertisement to young people.

Finally, mobile Internet video app designers can take advantage of our findings on viewing sources. Despite of various video discovery methods, the search and category page portals are two most popular viewing sources. A good search engine and an effective algorithm for video promotion to the category pages are therefore important. The designers should also consider the potential impact on geographic popularity imposed by viewing source.

VI. Related Work

The geographic features of video consumption have been studied recently to improve the quality of users' experience and the performance of systems. Cha *et al.* [10] examined the geographical locality of a IPTV system and found a large variation of interests on TV programs across geolocations. The geographic patterns of YouTube videos were examined in [4] [6] [12]. Brodersen *et al.* [4] analyzed the geographic popularity of YouTube videos. Huguenin *et al.* [6] examined the correlation between content locality and geographic locality. Scellato *et al.* [12] proposed to use the location information in Twitter to predict the geographic popularity of YouTube videos. We have compared the geographic popularity between YouTube and the PPTV mobile VoD systems throughout of our paper. The content property (*i.e.* user generated vs. commercially produced) and the viewing source (*i.e.* social network vs. mobile app portal) result in a different geographic patterns of videos.

Users in online social networks provide their location information Wittie *et al.* in [13] found that users in Facebook are likely to communicate with those within the same geographic region. The similar geographic locality of interests was also found in Twitter [11]. Backstrom *et al.* [3] found that the probability of friendship is inversely proportional to the distance at medium and long ranges. Kamath *et al.* [18] found the same temporal features of tweets in Twitter as that of videos in YouTube [4], since both tweets and videos greatly rely on social networks for content diffusion.

The user viewing behavior in mobile video systems have been studied in [8] [9]. In [8], authors studied the traffic patterns and user behavior for a mobile TV system. Li *et al.* [9] examined the impact of the connection type and the type of mobile device used on user behavior of accessing videos with mobile terminals. Finamore *et al.* [17] study the impact of devices (mobile device and PC) on user experiences in YouTube. Authors in [14] found that mobile apps show different geographic usage patterns, some are globally popular while some are mainly used by users in a few states in US.

VII. Conclusion

We make a first step towards understanding the geographic popularity patterns of a large-scale commercial mobile VoD system and the evolution trends over time. We surprisingly find a considerable fraction of videos attract views only from a few locations even in a commercial VoD system in which videos are professionally produced and aim at a global audience. More importantly, popular and non-popular videos exhibit distinct geographic popularity evolution trends. Non-popular videos attract most of their views from a set of a few locations, the diversity of which however varies greatly from day to day. On the other hand, popular videos' daily geographic popularity distributions closely follow the distribution of global traffic and remain stable. We also evaluate the impact of content type and viewing source from the mobile app. Finally, we discuss the implications of our findings on content delivery system, advertisement and mobile Internet video app design.

Acknowledgment

This work was supported by the National Basic Research Program of China with Grant 2012CB315801, the NSF of China (NSFC) with Grants 61133015 and 61272473, the National High-tech R&D Program of China with Grant 2013AA013501, and the CNGI-12-03-007 project. The work was also partially supported by the NSFC-ANR PFlower project 61061130562 and the Shanghai Science and Technology Committee with Grant 11DZ1500700.

References

[1] Cisco visual networking index: Global mobile data traffic forecast update, 2011-2016. Technical report, Cisco, 2012.
[2] PPTV mobile apps http://download.pptv.com, 2013.
[3] L. Backstrom, E. Sun, and C. Marlow. Find me if you can: improving geographical prediction with social and spatial proximity. In *Proceedings of WWW'10*, 2010.
[4] A. Brodersen, S. Scellato, and M. Wattenhofer. Youtube around the world: geographic popularity of videos. In *Proceedings of WWW'12*, 2012.
[5] J. Ding, L. Gravano, and N. Shivakumar. Computing geographical scopes of web resources. In *Proceedings of VLDB'00*, 2000.
[6] K. Huguenin, A.-M. Kermarrec, K. Kloudas, and F. Taïani. Content and Geographical Locality in User-Generated Content Sharing Systems. In *Proceedings of ACM NOSSDAV*, 2012.
[7] S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22, 1951.
[8] Y. Li, Y. Zhang, and R. Yuan. Measurement and analysis of a large scale commercial mobile internet tv system. In *Proceedings of IMC '11*, 2011.
[9] Z. Li, J. Lin, M.-I. Akodjenou, G. Xie, M. A. Kaafar, Y. Jin, and G. Peng. Watching videos from everywhere: a study of the pptv mobile vod system. In *Proceedings of IMC'12*, 2012.
[10] M. Cha, P. Rodriguez, J. Crowcroft, S. Moon, X. Amatriain. Watching television over an IP network In *Proceedings of IMC'08*, 2008
[11] T. Rodrigues, F. Benevenuto, M. Cha, K. Gummadi, and V. Almeida. On word-of-mouth based discovery of the web. In *Proceedings of IMC'11*, 2011.
[12] S. Scellato, C. Mascolo, M. Musolesi, and J. Crowcroft. Track globally, deliver locally: improving content delivery networks by tracking geographic social cascades. In *Proceedings of WWW'11*, 2011.
[13] M. P. Wittie, V. Pejovic, L. Deek, K. C. Almeroth, and B. Y. Zhao. Exploiting locality of interest in online social networks. In *Proceedings of CoNext'10*, 2010.
[14] Q. Xu, J. Erman, A. Gerber, Z. Mao, J. Pang, and S. Venkataraman. Identifying diverse usage behaviors of smartphone apps In *Proceedings of Imc '11*, 2011.
[15] R. Zhou, S. Khemmarat, and L. Gao. The impact of YouTube recommendation system on video views In *Proceedings of IMC'10*, 2010.
[16] H. Yin, X. Liu, T. Zhan, V. Sekar, F. Qiu, C. Lin, H. Zhang, B. Li. Design and deployment of a hybrid CDN-P2P system for live video streaming: experiences with LiveSky In *Proceedings of ACM MM'09*, 2009
[17] A. Finamore, M. Mellia, M. M. Munaf, R. Torres, S.G. Rao, YouTube everywhere: impact of device and infrastructure synergies on user experience In *Proceedings of IMC'11*, 2011
[18] K. Y. Kamath, J. Caverlee, K. Lee, and Z. Cheng. Spatio-temporal dynamics of online memes: a study of geo-tagged tweets In *Proceedings of WWW '13*, 2013