YouTube Around the World: Geographic Popularity of Videos

Anders Brodersen
Google
abrodersen@google.com

Salvatore Scellato*
University of Cambridge
ss824@cam.ac.uk

Mirjam Wattenhofer
Google
mirjam@google.com

ABSTRACT

One of the most popular user activities on the Web is watching videos. Services like YouTube, Vimeo, and Hulu host and stream millions of videos, providing content that is on par with TV [21]. While some of this content is popular all over the globe, some videos might be only watched in a confined, local region.

In this work we study the relationship between popularity and locality of online YouTube videos. We investigate whether YouTube videos exhibit geographic locality of interest, with views arising from a confined spatial area rather than from a global one. Our analysis is done on a corpus of more than 20 millions YouTube videos, uploaded over one year from different regions. We find that about 50% of the videos have more than 70% of their views in a single region. By relating locality to viralness we show that social sharing generally widens the geographic reach of a video. If, however, a video cannot carry its social impulse over to other means of discovery, it gets stuck in a more confined geographic region. Finally, we analyze how the geographic properties of a video's views evolve on a daily basis during its lifetime, providing new insights on how the geographic reach of a video changes as its popularity peaks and then fades away.

Our results demonstrate how, despite the global nature of the Web, online video consumption appears constrained by geographic locality of interest: this has a potential impact on a wide range of systems and applications, spanning from delivery networks to recommendation and discovery engines, providing new directions for future research.

Categories and Subject Descriptors

H.3.5 [**Information Storage and Retrieval**]: Online Information Services—*Web-based services*

General Terms

Measurement, Design

Keywords

Online Video Sharing, Geographic Popularity Analysis, Social Content Diffusion

*Salvatore Scellato conducted this research during an internship at Google.

WWW 2012, April 16-20, 2012, Lyon, France
ACM 978-1-4503-1229-5/12/04.

1. INTRODUCTION

Historically, most of the media content was distributed via media organizations that often segmented users in regional markets, releasing new content in a controlled way. Hence, video popularity was seldom a global phenomenon, as users could not access the same content all over the world. With the advent of online video sharing platforms like YouTube such regional barriers were largely demolished, making most content items accessible from all over the globe.

With more than 3 billion videos viewed everyday and having more than 70% of its traffic coming from outside the US [20] YouTube is an ideal data source to authoritatively answer the question whether, despite global availability, most videos enjoy only local popularity and, if yes, how.

To understand the rules that govern the geographic reach of a video, it is important to first understand the general access pattern of YouTube videos. It was shown by Cha et al. [5] that video popularity on YouTube exhibits a "long-tail" behavior: some videos are able to accumulate hundreds of millions of views, whereas the vast majority can only attract a few. This unbalanced skew of video popularity can be exploited to serve the huge interest for a potentially unlimited number of relatively unpopular items [1]: users are able to discover and enjoy millions of videos about niche topics they are interested in, even though each individual video might not accrue a large number of overall views. How video popularity relates to the global reach of a video, and whether unpopular videos still enjoy high popularity *in a certain geographic area* is an open question we will address in this paper.

Geographic locality of interest.

Several factors may give rise to geographic locality of interest in online videos. For instance, topics like sports, politics, and news, tend to have a spatial focus of interest [8, 2], and thus *geographic relevance* is a powerful factor impacting video popularity. Another factor is *geographic proximity* between users, since web content tends to spread through word-of-mouth across social connection [15]. Since users close to each other tend to exhibit similarities in language and culture [13, 3], spatial closeness may also constrain the propagation of video web links over online social networking platforms: as recently studied, a large fraction of such cascading steps happen at relatively short-range spatial distances [16, 15]. Consequently, the geographic scope of a video might well be constrained to web users in a limited geographic region, with important consequences to how relevant it might be to users in that region [8].

Besides its impact on users, geographic locality of interest has also an impact on systems and infrastructures. In fact, skewed popularity of content items positively impacts caching mechanisms, which are able to serve a large fraction of requests by storing only a handful of the most popular items [18]. When popularity is geographically scoped, distributed caching mechanisms can be devised to optimize performance: spatial locality of interest potentially represents a great advantage for modern geographically distributed content delivery networks and data centers. These large-scale infrastructures distribute their content from storage servers replicated across multiple locations over the planet, aiming to maximize the overall efficiency [12]. Whenever items appear predominantly requested from a given geographic area it becomes possible to devise mechanisms that improve delivery performance: by exploiting the simple fact that future requests of a given item are more likely to be generated at a close geographic distance from the previous accesses, spatially distributed caching servers can achieve high hit ratios[19, 16]. These aspects become even more important when considering how both social mechanisms and geographic location affect how content is consumed on the Web[16, 14].

Finally, online video streaming is forecasted to account for 50% of consumer Internet traffic by 2012 [6]. Since YouTube constitutes one of the largest sources of this traffic, understanding how and where users watch YouTube videos becomes of paramount importance and provides insights useful across several domains, ranging from building predictive models of user interest to designing recommending systems.

Our approach.

The question we address in this work is whether a *geographic locality of interest* is effectively experienced by YouTube videos and, if so, *how to measure and understand it.* In particular, we aim to answer these research questions:

- do YouTube videos experience geographic locality of interest or, rather, a uniform view popularity across the world?
- how are social and geographic factors influencing the spatial properties of YouTube video traffic?
- do YouTube videos experience a uniform geographic interest over their lifetime or, instead, do they exhibit distinctive patterns in their temporal evolution?

In order to illustrate these issues and try to answer these questions, our approach focuses on directly studying how YouTube videos enjoy views from different parts of the world over their lifetime. In addition, we study how these views might be generated by social rather than non-social mechanisms and how other characteristics such as the type of video content and the main geographic area of interest affect the geographic provenance of video views. We analyze a large corpus of YouTube videos: since each single viewing event can be assigned to a geographic region, we are able to define concepts such as the *focus location* of a video, its *view focus* and its *view entropy*, a measure of the dispersion of its audience. By adopting these quantitative concepts we are able to track and compare how a video exhibits different geographic patterns of interest and to numerically assess the impact of other properties on its geographic audience (Section 2).

Our contributions.

With our analysis we uncover some surprising facts about the geographic properties of YouTube video popularity, while at the same time we are able to confirm how YouTube videos enjoy a strong geographic locality of interest:

- **YouTube videos are local**: we find that *about 50% of videos have more than 70% of their total views in a single country* (which holds even for popular videos) (Section 3);
- **Viral spreading traps videos**: we analyze that the impact of social sharing on the geographic properties of video views is surprisingly non-trivial. As a video receives a larger fraction of views through social mechanisms its geographic audience widens, but *when the fraction of socially-generated views grows larger than 20% the videos experience a more focused popularity in fewer regions* (Section 4);
- **Popularity expands and withdraws back**: we discuss that video popularity evolves over time, exhibiting a peak of interest which then fades away, *video views immediately grow in the focus location and only then they expand across other regions*, withdrawing back to the focus location after the peak (Section 5).

2. METHODOLOGY

In this work we study a corpus of more than 20 million YouTube videos randomly selected from the set of all videos uploaded to YouTube between September 2010 and August 2011. Adopting a year-long sampling period allows us to avoid the average seasonalities exhibited by user behavior over the year, albeit YouTube traffic was steadily increasing over this period.

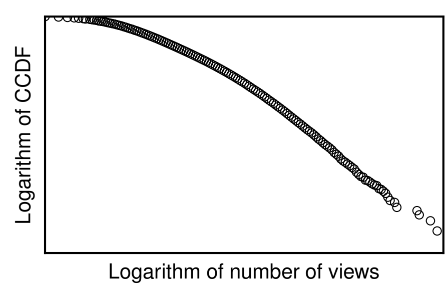
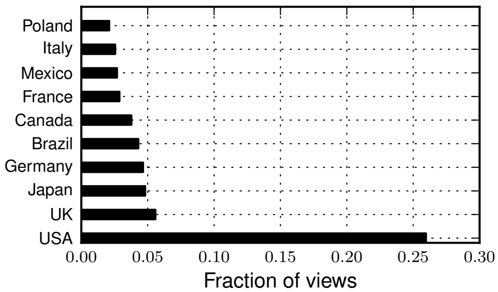


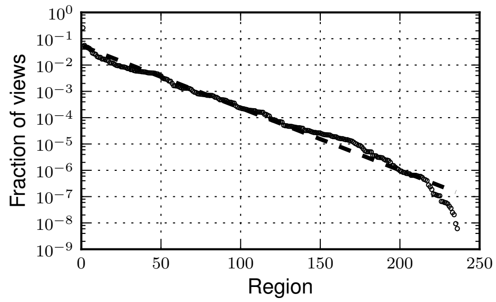
Figure 1: Complementary Cumulative Distribution Function of the number of lifetime views V_i for each video i: the most viewed videos contribute to the heavy tail of the distribution.

2.1 Data description

We have access to the number of daily views coming from different geographic regions for each video: these regions represent national countries or other political and geographic entities. Overall, there are about 250 different such regions, including official states and minor territories. The data available for each video include the video category and the breakdown of views by referrer, that is, by the website the user arrived from. We classify these referrer sources as



(a)



(b)

Figure 2: Fraction of video views generated across different regions: top 10 regions (a) and as a function of a region's rank (b), where the rank is computed by views from that region. Only a quarter of total traffic is generated within the USA, while the fraction of views decreases exponentially for other regions.

either *social* or *non-social* [4]: we consider a referrer *social* only in the cases when the user directly typed the URL into the browser, viewed the video embedded in a third-party website or clicked on a link on an external website. All the other sources, mainly related to web search or YouTube internal navigation, are considered *non-social*.

2.2 Notation

Given a video i, we represent the distribution of its views on day t across M different regions $(r_1, \dots, r_M)$ as a vector $(v_{i1}^t, v_{i2}^t, \dots, v_{iM}^t)$ and we denote its total number of views on day t as $V_i^t = \sum_{k=1}^{M} v_{ik}^t$. We denote with t_i the day video i was uploaded and we define lifetime metrics for each video by defining $v_{ik} = \sum_{t \geq t_i} v_{ik}^t$ as the lifetime views received by video i in region r_k, with $V_i = \sum_{k=1}^{M} v_{ik}$ being the total number of views received by video i over its entire lifetime.

We define s_i^t as the number of views received from social referrers by video i on day t and $s_i = \sum_{t \geq t_i} s_i^t$ as the total number of these views. Thus, we define the daily social ratio as $S_i^t = \frac{s_i^t}{V_i^t}$ and the lifetime social ratio as $S_i = \frac{s_i}{V_i}$.

2.3 Basic properties

The set of videos under analysis encompasses a wide range of popularity levels, from videos with a handful of views to videos with millions of hits. The probability distribution of lifetime video views V_i, depicted in Figure 1, exhibits a typical heavy-tail behavior, with a few, extremely popular videos accruing the main numbers of requests. At the same time, the vast majority of videos has only few views: 50% of YouTube videos have less than 50 views over the period under analysis.

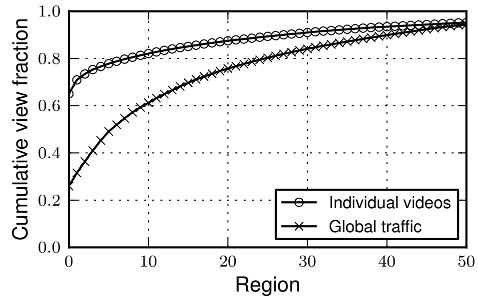


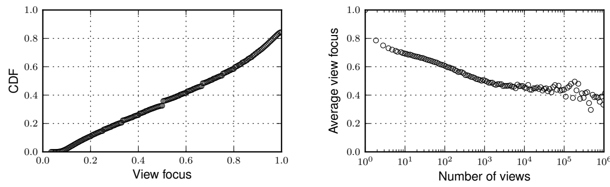
Figure 3: Average cumulative fraction of views in the top k regions as a function of the region rank k, for individual videos and for the global YouTube traffic. For a given video, the top 10 regions mount up to, on average, more than 80% of the total number of views.

Analyzing how YouTube views are distributed across the world confirms that the majority of YouTube traffic does not come from the USA [20]: as shown in Figure 2(a), only 26% of YouTube views are generated in the USA, with other regions generating at maximum about 5% of views each. More in detail, the global distribution in Figure 2(b) shows how regions exhibit an approximately exponentially decreasing fraction of views when they are ranked. This is largely due to the fact that many of these regions are fairly small and/or have a comparatively small Internet population. Thus, the views received by YouTube videos are likely to concentrate in a handful of key regions which contain a large amount of the world population of web users.

However, this concentration effect for individual videos is much stronger than the global distribution suggests. In fact, *on average YouTube videos acquire a large fraction of their views from only a very small number of regions*: as displayed in Figure 3, whereas the top 10 countries generate only about 60% of YouTube global views, for individual videos the average value is higher than 80%. Hence, single videos seem to enjoy a vast share of their popularity in a few, key countries, exhibiting a highly localized interest.

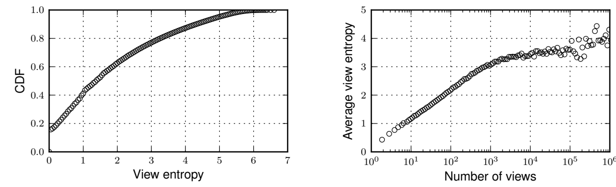
2.4 Locality measures

In order to quantitatively study whether a YouTube video receives views from a global rather than from a local audience we exploit the segmentation of views across different regions: our aim is to define some measures to understand whether videos exhibit locality of interest effects, allowing us to investigate the potential causes of such behavior. The main idea behind these measures is that a video enjoying a highly localized popularity would likely exhibit a non-uniform distribution of its views across different regions, with a large fraction of them in only a few regions. On the other hand, videos with a global pattern of popularity are likely to have more uniform distribution of their views across the regions.



(a) (b)

Figure 4: Cumulative Distribution Function (CDF) of view focus H_i (a) and average view focus as a function of the lifetime video views V_i for each video.



(a) (b)

Figure 5: Cumulative Distribution Function (CDF) of view entropy H_i (a) and average view entropy as a function of the lifetime video views V_i for each video.

Previous attempts to define the geographic scope of web resources have focused on measuring the *peak intensity* and the *uniformity* of the interest that such resources trigger over a set of geographically-defined regions [8]. We adopt a similar methodology, defining two key measures for each YouTube video: the **view focus** and the **view entropy**.

The *view focus* F_i of video i is defined as

$$F_i = \frac{1}{V_i} \max_k v_{ik} \tag{1}$$

The view focus is the highest fraction of views that video i has received in a single region over its entire lifetime. As a consequence, we define L_i as the *focus location* of video i, that is, the region where video i has its view focus, randomly breaking ties between regions. In a similar way, we define the *daily view focus* F_i^t considering only the daily views of video i on day t.

The *view entropy* H_i of video i is defined as

$$H_i = -\sum_{k=1}^{M} \frac{v_{ik}}{V_i} \log_2 \frac{v_{ik}}{V_i} \tag{2}$$

where the sum is running only over regions for which $v_{ik} \neq 0$. As its name suggests, the view entropy is effectively the information entropy of the distribution of views of video i over the different regions. Hence, higher values of view entropy denote videos whose views are spread more uniformly across several regions, while lower entropy values signal video views more focused in less regions. In the rest of this paper, we always express numerical values of view entropy in bits. Again, we define the *daily view entropy* H_i^t considering only the daily views of video i on day t.

3. LOCALITY OF INTEREST

We start our analysis by studying the locality measures previously defined, discussing the overall predominance of locality of interest in YouTube videos. Then, we investigate how videos belonging to different categories, and popular in different regions, present diverse patterns of geographic popularity.

3.1 Measuring locality of interest

It is worth highlighting that if every single video experienced a distribution of views across different regions as the distribution of global YouTube traffic, depicted in Figure 2, then on average videos would exhibit a view focus of about 0.26 (that is the fraction of views coming from the USA) and a view entropy of about 5 bits. Instead, individual videos exhibit not only a wide range of values of these metrics, but also values that signal a more localized interest from users in a small subset of regions.

The distribution of view focus in Figure 4(a) shows hows videos exhibit the whole range of values, with the median value close to 0.7. In particular, there are about 40% of YouTube videos that enjoy at least 80% of their views in a single region. This is already strong evidence supporting the claim that videos tend to become popular in a locally confined area, rather than in a globally wide region. Furthermore, as videos accumulate more views, they tend to be watched in a more disparate set of regions: as shown in Figure 4(b), the average view focus decreases linearly as the order of magnitude of the number of views grows. Nonetheless, even videos with more than 1,000 views, which may be deemed already fairly popular, exhibit a steady value of about 0.4. This is still higher than one would expect by observing the distribution of global YouTube traffic in Figure 2: in fact, no region exhibits more than 26% of the overall number of views.

A related trend can be observed when considering the view entropy, which captures how video views are uniformly spread over different regions. Again, videos have values much lower than the entropy of the global view distribution: as seen in the distribution in Figure 5(a), videos have a median view entropy of 1.5 bits, with about 40% of them with view entropy lower than 1 bit. On the contrary, view entropy grows with the number of views: as the order of magnitude of the number of views grows, view entropy grows linearly. However, for videos with more than 1,000 views this growth is much slower: view entropy values reach a plateau at about 3.5 bits.

It is important to consider that only 3% of videos in our sample have only 1 view, which directly results into a maximum view focus of 1.0 and a minimum view entropy of 0. Nonetheless, there is a significant fraction of videos, about 18% of them, that show such extreme locality of interest when considering both view focus and view entropy.

3.2 Impact of video category

Since YouTube videos cover several different topics, which might have a more global or local geographic appeal, it appears natural to study the locality of interest exhibited by different video categories. We report the average view focus and the average view entropy to each category in Table 1.

All the categories have a similar median number of views. However, their locality metrics span a wide range. In particular, the categories "Category 13", "Category 14" and "Category 4" enjoy the highest levels of view entropy and the lowest levels of view focus, denoting an average global ap-

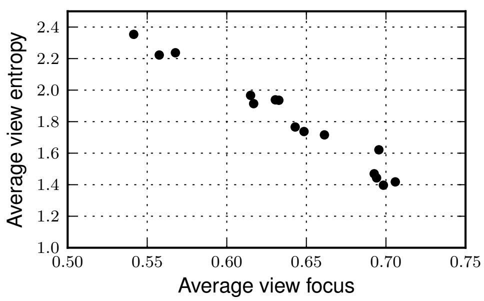


Figure 6: Scatter plot of the average view focus and the average view entropy of videos belonging to the same category.

Category	$\langle F_i \rangle$	$\langle H_i \rangle$
Category 1	0.69	1.39
Category 2	0.63	1.93
Category 3	0.61	1.91
Category 4	0.55	**2.22**
Category 5	0.70	**1.41**
Category 6	0.69	1.62
Category 7	0.61	1.96
Category 8	0.64	1.76
Category 9	0.63	1.93
Category 10	0.64	1.73
Category 11	0.69	1.44
Category 12	0.69	1.46
Category 13	0.54	**2.35**
Category 14	0.56	**2.23**
Category 15	0.66	1.71

Table 1: Properties of videos belonging to different categories: average view focus and average view entropy.

peal of those videos. On the other hand, videos belonging to the "Category 5" category display the most local measures, showing that these videos are more likely to have a geographically limited audience because of language and cultural constraints. The correlation between view focus and view entropy for different categories is depicted in Figure 6: high values of view entropy tend to correlate with lower values of view focus. This confirms how the topic of a video is strongly affecting how users from a wide geographic area, rather than a more focused one, are interested in it.

3.3 Impact of video location

Another important aspect which might be affecting the spatial properties of a video's views is the location a video is related to. It is not hard to imagine how items related to some regions are more accessible by users elsewhere, because of similar language, culture or interests. Even though it is often difficult to assign a video to a certain region, a reasonable and unambiguous proxy is to adopt the focus location of a video as the region that video is related to and study the properties of videos that have the same focus location, comparing them across different regions.

As a result, in Table 2 we present the top 10 regions by number of videos, together with their median number of views, their average view focus and their average view entropy.

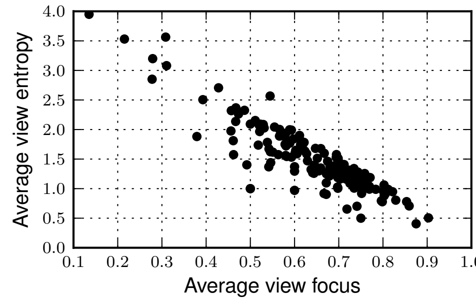


Figure 7: Scatter plot of the average view focus and the average view entropy of videos with the same focus location.

Region	Videos	Median V_i	$\langle F_i \rangle$	$\langle H_i \rangle$
USA	37.7 %	0.513	0.75	1.24
Brazil	6.6 %	0.497	**0.90**	0.49
UK	4.9 %	0.460	0.69	1.38
Germany	4.1 %	0.937	0.73	1.30
Japan	3.2 %	1.000	**0.85**	0.71
Spain	2.9 %	0.534	0.75	1.12
France	2.5 %	0.630	0.73	1.28
Mexico	2.5 %	0.647	0.70	1.34
Canada	2.5 %	0.323	0.72	1.08
Italy	2.4 %	0.859	0.81	0.96

Table 2: Properties of videos with their focus location in different regions: fraction of videos, median number of views (divided by the median number of views for Japan), average view focus and average view entropy.

More than one third of all videos have their focus location in the USA, while other countries have less than 5% of videos, with the exception of Brazil. This reflects the overall predominance of YouTube traffic generated in the USA, already observed in Section 2. More interestingly, some countries such as Brazil and Japan exhibit high values of view focus: on average, videos whose focus location is Japan or Brazil have, respectively, 90% or 85% of their views in their focus location. These values denote a highly local interest for videos related to those regions, as supported also by the low values of view entropy in these same regions.

On the other hand, places such as the UK, Mexico and Canada exhibit more global spatial measures, with lower view focus and higher view entropy. A potential explanation for these values is that such regions might enjoy traffic coming from the USA, since there is a likely vast interest overlap among users across these places. Overall, the location a video is related to is strongly affecting where its views are coming from, with some regions much more likely to exhibit highly local content and others more likely to have content with a wider geographic appeal. Again, the overall interest, measured by the number of views received by videos in different regions, does not seem to affect the spatial properties of video views. Instead, there is a strong correlation between lower values of view entropy and higher values of view focus: as shown in the scatter plot of Figure 7, this relationship seems to hold across all the regions considered in our analysis.

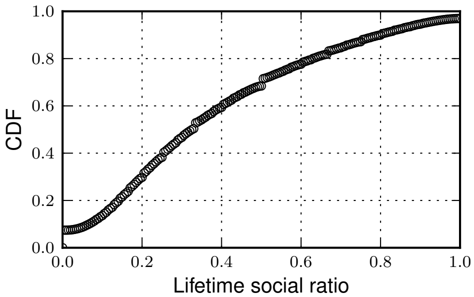


Figure 8: Cumulative Distribution Function of the lifetime social ratio S_i for all videos.

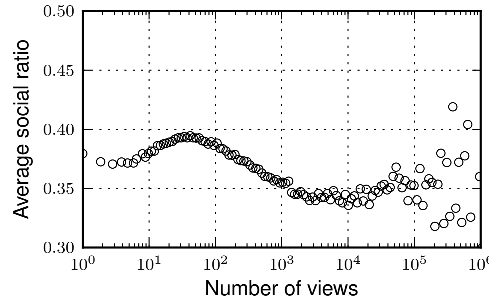


Figure 9: Average lifetime social ratio S_i as a function of the number of lifetime views V_i for each videos.

4. SOCIAL FACTORS

Among the factors driving the success of YouTube there is the social diffusion of individual videos from person to person across online and offline media. Given the sheer amount of available videos, it is often difficult for users to find and discover interesting content: thus, relying on suggestions coming from friends is a popular and effective way to choose what to watch. At the same time, some videos enjoy a viral spreading thanks to social cascading, resulting in high levels of overall popularity [4, 9].

In this section our goal is to understand whether social sharing is affecting not only video popularity, but also the geographic properties of video popularity. More in detail, since social interaction among web users appears affected by geographic distance between them, with individuals more likely to interact with closer users than with users further away [3, 17], our question is whether higher levels of social sharing experienced by YouTube videos result in a more local geographic pattern of popularity.

4.1 Social ratio

As discussed in Section 2, we can classify each view event as social or non-social by considering how the user arrived at the video player, defining the lifetime social ratio of a video as the fraction of lifetime views coming from social sources. The distribution of the lifetime social ratio S_i, presented in Figure 8, highlights how YouTube videos enjoy a wide range of levels of social sharing, with an average value of social ratio of about 0.37.

However, the amount of social sharing experienced by YouTube videos is different for videos with different number of lifetime views. As shown in Figure 9, the lifetime social ratio increases with the number of views for videos with less than 50 views, while it then decreases steadily for videos with more views. It appears that the impact of social sharing is higher for videos with few views, while when videos enjoy higher popularity levels then social sharing becomes less prominent.

4.2 Impact of social sharing

Since the amount of social sharing that a video enjoys is related to its overall popularity, our aim is to investigate the impact of social sharing on the geographic popularity. In detail, our aim is to understand whether higher levels of social sharing results in a more local, or global, video geographic popularity. To our surprise, this effect is not simply linear: videos with a lifetime social ratio close to 0.2, thus with about 1 out of 5 views generated by social mechanisms, exhibit more global popularity than other videos.

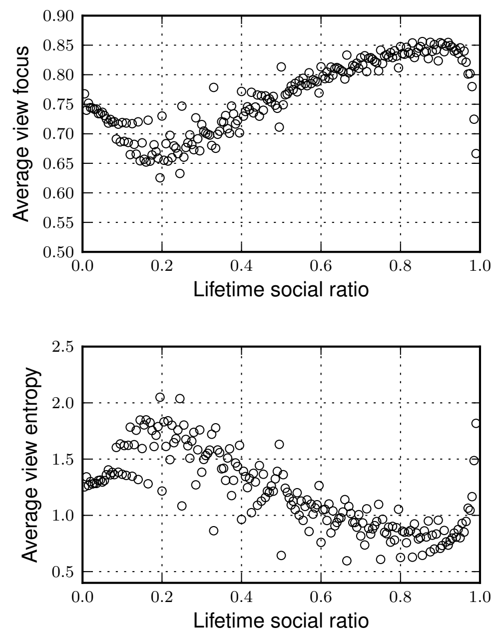


Figure 10: Average view focus (a) and average view entropy (b) as a function of the lifetime social ratio.

In fact, as displayed in Figure 10(a), the average view focus decreases as the social ratio increases from 0 to 0.2, while then the view focus goes up again for higher levels of social sharing. This peculiar behavior might be due to the fact that a low level of social sharing helps the video spreading across several regions. Instead, when social views are predominant, it is more likely that social diffusion, constrained by geographic distance, happens within the boundaries of a single geographic region. A similar trend is observed for the view entropy in Figure 10(b): as the level of social ratio approaches 0.2 the view entropy is higher, denoting a widespread attention received by the video from different geographic regions. Higher levels of social sharing, as well as lower levels, reduce the geographic audience of the video.

However, videos that enjoy different levels of overall popularity might exhibit different behavior than the overall trend,

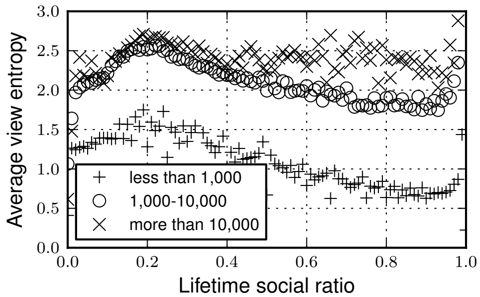


Figure 11: Average view entropy as a function of the fraction of social views for videos with less than 1,000 views, with a number of views between 1,000 and 10,000 and with more than 10,0000 views (c).

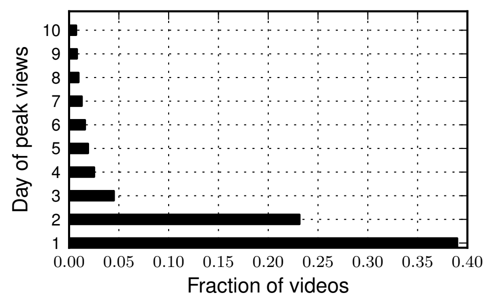


Figure 12: Fraction of videos with their peak daily views in different days of their lifetime. Only videos with more than 100 views in their first 30 days are considered.

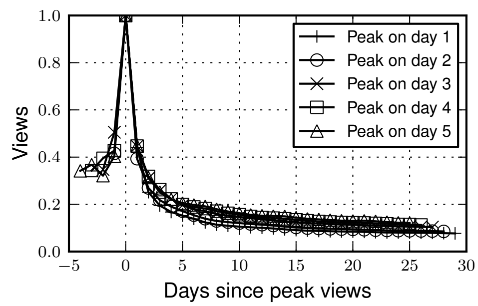


Figure 13: Average normalized number of daily views for videos peaking on different days of their lifetime. Only videos with more than 100 views in their first 30 days are considered.

which is clearly dominated by the vast majority of videos that have only a handful of views. In order to understand the impact of social sharing for videos with higher number of views we segment all videos into 3 sets: videos with less than 1,000 lifetime views, videos with a number of lifetime views between 1,000 and 10,000 and videos with more than 10,000 lifetime views. The effect of social sharing for these three groups of videos is relatively different across these three categories: as displayed in Figure 11, whereas videos with less than 1,000 views have an average view entropy which still follows the overall trend of Figure 10(b), with a maximum value for values of social ratio of about 0.2, more popular videos with 1,000 to 10,000 views have a similar pattern but with higher values and then videos with more than 10,000 views exhibit less dependency on the amount of social sharing, with an average view entropy uncorrelated to the social ratio.

5. TEMPORAL EVOLUTION

After discussing how video properties and social sharing affect spatial popularity, we shift our attention to the temporal evolution of the geographic patterns of video views. More in detail, we are interested in studying whether videos exhibit a steady and uniform level of locality of interest across their lifetime or if, instead, their geographic audience changes over time. Finally, we will discuss the relationship between the amount of social sharing enjoyed by a video and the temporal evolution of its geographic audience.

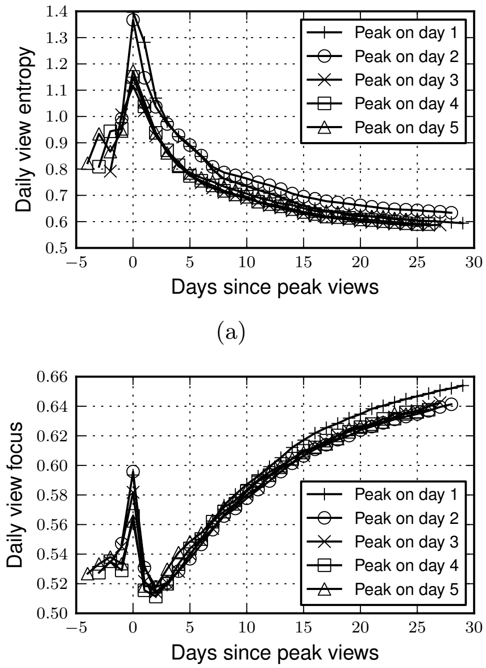
Hence, in this section we take advantage of the daily temporal granularity of video views to study how the number of views, and the related view focus and view entropy, evolve over time. Videos with an extremely low number of views do not provide enough data to study their temporal evolution: thus, all results in this section consider only videos that receive at least 100 views within their first 30 days of viewing.

5.1 Video Views Growth

In order to aggregate and study videos that in reality are uploaded and watched during different temporal periods, we align their viewing history in time. We exploit the fact that videos often experience a peak in their number of views when they are featured on popular websites or when they become viral and spread on online social networks. It has also been suggested that the peculiar way in which the burst of views evolves over time can be exploited to classify the popularity of a given video [7]. Hence, we choose the day when their daily views peak as day 0, following a methodology already successfully adopted for similar purposes [4].

As presented in Figure 12, the videos in our sample are more likely to peak on their first viewing day, with about 38% of them doing so. The number of videos that peak on successive days is increasingly smaller, with only 22% of videos peaking on day 2 and 5% on day 3. Even though videos do peak on different days of their lifetime, once their daily views are aligned there is no significant difference in their temporal evolution: after the peak the number of daily views falls back to about 20%-10% of the peak value, regardless of the peaking day.

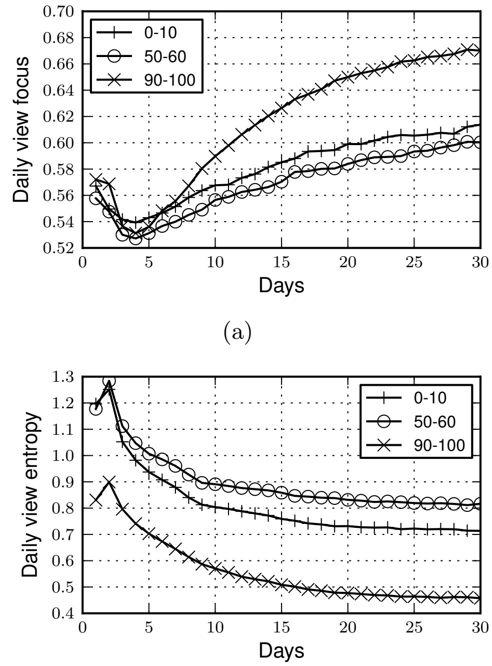
While this pattern may be completely different for individual videos, especially in the cases where they are prominently featured and promoted, the overall trend appears robust. Nonetheless, other studies have discussed how the peculiar pattern exhibited by the time series of daily views, particularly *before* and *after* the peak, can be used to classify popular videos [7]. This would suggest that our overall temporal pattern is due to the fact that our much larger dataset does not focus only on popular videos, but on a larger representative sample of YouTube videos. Thus, individual time series of hugely popular videos, which might show the different temporal patterns observed in other investigations, do not represent the temporal evolution most commonly experienced by YouTube videos.



(a)

(b)

Figure 14: Average daily view entropy (a) and daily view focus (b) over a video lifetime, for videos peaking in different days of their lifetime. Only videos with more than 100 views in their first 30 days are considered.



(a)

(b)

Figure 15: Average daily view focus (a) and average daily view entropy (b) across different days of a video lifetime for three video segments with different fraction of social views. Only videos with more than 100 views in their first 30 days are considered.

5.2 Evolution of local interest

As the number of daily views peaks and then falls down, the spatial properties of these views are likely to change as well. As we have already discussed in Section 3, videos with more lifetime views exhibit a wider spatial audience, with higher values of view entropy and lower values of view focus. Thus, our aim is to understand whether videos show a similar behavior when their daily number of views peaks.

To our surprise, we find that while the daily view entropy increases when the number of daily views peaks, the daily view focus increases as well, as shown in Figure 14. Even when videos peak on different days of their lifetime they tend to exhibit this pattern, the only difference being that videos peaking on later days tend to have a higher daily view entropy on the peak day. More interestingly, while daily view entropy slowly decreases after the peak day, the daily view focus has a sudden drop in the following days after the peak and then it steadily increases again over the next days.

These values suggests a scenario where a video suddenly experiences a large fraction of views coming from its focus location, while also receiving additional views coming from other regions. This has the effect of pushing up both the view focus and the view entropy. Immediately after the peak the fraction of views coming from the focus location goes down, as interest shifts to other regions. Finally, view traffic shifts back to the focus location, since view entropy decreases and view focus increases.

This "spray-and-diffuse" pattern indicates how, on average, a video tends to become popular and to peak in its own focus location, and only then the interest generated moves to other regions. At the end, as popularity fades away video views retreat back to the focus location.

5.3 Effect of social sharing

Finally, we analyze the impact of social sharing on the temporal evolution of the spatial properties. Again we time-align the viewing history of each video, so that the view peak is on day 0, and we consider the evolution of the daily view measures, view focus and view entropy, averaged over all the videos. However, now we consider the fraction of social views videos receive during their first 30 days and then we divide them in ten quantiles, thus with approximately the same number of videos in each group. The first quantile contains videos with a fraction of social views lower than 11.8%, whereas the last quantile has videos with at least 74.3% social views.

In Figure 15(a) we show that the main difference is how videos that enjoy higher levels of socially-driven traffic tend to have a higher long-term value of view focus, whereas on the peak days the different classes behave similarly. In the case of view entropy, depicted in Figure 15(b), even on the peak day videos with higher fraction of social views have lower view entropy and, at the same time, the long-term trends are different for different segments of videos, with high "socialness" correlated with lower view entropy.

These patterns suggest that the effect of high levels of social sharing on the temporal evolution is that on the peak day video views are more focused in a single region than on several ones. Moreover, the long-term trend after the peak is directly related to the amount of social sharing, with more social videos enjoying more locally focused popularity. Overall, these results confirm our previous analysis of the impact of social sharing even on the temporal daily metrics of YouTube videos.

6. IMPLICATIONS

The main result of this work is that even though web video services address a global audience, much differently than how standard media distribution platforms segmented the world in regional markets, yet online videos experience patterns of popularity that appear strongly constrained by geographic locality of interest, with user interest predominantly coming from few key regions for a vast majority of items. We believe that even though our analysis has focused only on YouTube, its wide popularity and its massive user base allow us to gain insight not only on user behavior on YouTube itself, but on potential user behavior on other similar online media platforms. Furthermore, we see a great potential to exploit the geographic properties of video popularity on online services.

An important application of our findings is in personalization: tailoring video content presented to users according to their geographic whereabouts appears as a viable way of engaging them more. This strategy has the potential to go beyond simple global popularity, in order to consider items that might not be noteworthy *per se* but that could become so for users belonging to a certain geographically-defined community. This could affect a wide set of services, ranging from search to discovery tool and recommendation engines.

The surprising effect of social sharing on the spatial properties of a video's audience suggest that the relationship between social diffusion mechanisms and the popularity of content is more complicated than it would appear. In detail, higher levels of social sharing appear to constrain, rather than widely promote, the diffusion of a piece of content on a more focused and less diverse set of geographic regions. This finding appears in agreement with many recent results on the spatial properties of social networks [3, 17], which have found how geographic distance affects social interaction on online social networks, acting as a powerful proxy for culture and language differences. Yet, the causality link between social sharing and geographic popularity is to be further studied.

At the same time, social diffusion mechanisms have been studied on different online platforms, investigating how information spreads by cascading through social links and how this phenomenon can be modeled or predicted. Our results opens new directions for the analysis of such social diffusion mechanisms on online services: geographic and spatial factors could be taken into account to model how users choose what content to share and which friends they want to share it with. In addition, the interaction between geographic distance, social diffusion and online popularity could lead to better classification and predictive models for online videos and online news.

At the same time, spatial locality of interest for online content has the capability to influence large-scale storage and delivery systems. For instance, as a majority of content items enjoy only local popularity, modern geographically distributed content delivery networks can preferentially cache different items in different geographic regions, exploiting how both social and spatial factors limit the diffusion of online items beyond certain geographic scales. In an era where billions of videos are streamed every day on the global Internet, the prospect of a geographically diverse and spatially-scoped Web could effectively ease the load that the skyrocketing amount of online video traffic imposes on the networking infrastructure.

7. RELATED WORKS

There are two main research areas related to our work: the analysis of the properties of online video services and the investigation of the spatial properties shaping user behavior on online services.

Online videos.

Research efforts have focused on understanding the properties of YouTube videos and how their popularity can be modeled, classified and predicted. One of the first works on YouTube focused on the popularity life-cycle of videos and on the statistical properties of user requests, discussing the impact of long-tail popularity on video discovery and content delivery infrastructure design [5]. The temporal dynamics of video popularity has been also investigated, studying how it evolves over time to build predictive models [7]: this has sparked interest about the *viralness* of YouTube videos, where individual video items become wildly popular thanks to social diffusion mechanisms, even though a thorough understanding of such phenomenon is still to be achieved [9, 4]. Our investigation is different than this body of works since we study an entire sample of YouTube videos rather than mainly popular ones; in addition, we address a different facet of YouTube video popularity, focusing on the novel aspect of its spatial properties in relation to video properties, amount of social sharing and temporal evolution.

Other researchers have measured how users access YouTube videos by monitoring local networks and studying how local popularity does not correlate with global popularity and how many videos enjoy high levels of interest even within a confined user population, suggesting how this can be exploited to provide local caching and proxying [10, 22]. We extend this set of results by discovering similar patterns of local popularity across the entire range of YouTube videos, confirming at a larger scale those results and building up on them to provide new insight.

Spatial properties of user behavior.

The effect of spatial distance on online user behavior has only been recently addressed, as more information about the location of users and of resources has become available, thanks to new location-sensing technologies. Nonetheless, one of the first studies on the geographic scope of web resource predates such technologies and discusses methods to compute the interest and the spatial spread of web pages [8]. Another work focuses on studying search engine queries to estimate their geographic center of interest and their spatial dispersion in the USA [2]. We adopt concepts already put forward by these works, namely studying the spatial distribution of item popularity by considering a focus location and the dispersion around it. However, we adapt these concepts to the YouTube scenario by defining the view focus and the view entropy measures across world regions.

Another thread of research has been investigating the spatial properties of user interaction on online social services. It has been found how the probability of having a social connection between two individuals decreases as an inverse power of their geographic distance, even though the exact relationship is still under discussion [13, 11, 3]. Furthermore, users exhibit heterogeneous socio-spatial properties, with geographic distance becoming less constraining as users acquire more online social ties [17]. While these results focus on the structural properties of online social networks, it has also been discussed how social diffusion mechanisms are constrained by spatial distance, with information propagation preferentially occurring over short-range social ties [16]. In particular, a recent study found how propagation of YouTube URLs on Twitter is widespread and preferentially taking place between geographically close users [15]. Our findings confirm these results, as we provide new evidence that connects social mechanisms with spatial distance, namely the fact that higher levels of social sharing tend to limit the geographic spread of video popularity, confirming the effect of geographic properties on online user behavior.

8. CONCLUSIONS AND FUTURE WORK

We have seen that YouTube videos exhibit strong geographic locality of interest by studying a large corpus of video items and using new measures to quantify their popularity distribution across different geographic regions. We have discussed the impact that video properties have on these measures, together with studying how social sharing affects the spatial popularity of a video and how these measures evolve over the lifetime of the video.

There are a number of directions that appear worth to be pursued to extend these results. Predictive and classification models for video popularity could take spatial properties into account to refine their performance, while search and recommendation engines may exploit these spatial measures to tailor their results and adapt them to the user's geographic location. Finally, availability of finer-grained geographic information about video items and video views would make spatial probabilistic models possible, exploring popularity pattern at even smaller geographic scales.

9. REFERENCES

[1] Chris Anderson. *The Long Tail: Why the Future of Business Is Selling Less of More*. Hyperion, 2006.

[2] Lars Backstrom, Jon Kleinberg, Ravi Kumar, and Jasmine Novak. Spatial variation in search engine queries. In *Proceeding of WWW'08*, pages 357–366, New York, NY, USA, 2008. ACM.

[3] Lars Backstrom, Eric Sun, and Cameron Marlow. Find me if you can: improving geographical prediction with social and spatial proximity. In *Proceedings of WWW'10*, pages 61–70, New York, NY, USA, 2010. ACM.

[4] Tom Broxton, Yannet Interian, Jon Vaver, and Mirjam Wattenhofer. Catching a viral video. In *ICDM'10 Workshop*, pages 296–304, Los Alamitos, CA, USA, 2010. IEEE Computer Society.

[5] Meeyoung Cha, Haewoon Kwak, Pablo Rodriguez, Yong-Yeol Ahn, and Sue Moon. I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In *Proceedings of IMC'07*, pages 1–14, New York, NY, USA, 2007. ACM.

[6] Cisco Systems Inc. Entering the Zettabyte Era. `http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/VNI_Hyperconnectivity_WP.html`. Accessed Sep. 2011.

[7] Riley Crane and Didier Sornette. Robust dynamic classes revealed by measuring the response function of a social system. *PNAS*, 105(41):15649–15653, 2008.

[8] Junyan Ding, Luis Gravano, and Narayanan Shivakumar. Computing geographical scopes of web resources. In *Proceedings of VLDB'00*, pages 545–556, San Francisco, CA, USA, 2000. Morgan Kaufmann Publishers Inc.

[9] Flavio Figueiredo, Fabrício Benevenuto, and Jussara M. Almeida. The tube over time: characterizing popularity growth of youtube videos. In *Proceedings of WSDM'11*, pages 745–754, New York, NY, USA, 2011. ACM.

[10] Phillipa Gill, Martin Arlitt, Zongpeng Li, and Anirban Mahanti. Youtube traffic characterization: a view from the edge. In *Proceedings of IMC'07*, pages 15–28, New York, NY, USA, 2007. ACM.

[11] Renaud Lambiotte, Vincent D Blondel, Cristobald De Kerchove, Etienne Huens, Christophe Prieur, Zbigniew Smoreda, and Paul Van Dooren. Geographical dispersal of mobile communication networks. *Physica A*, 387(21):17, 2008.

[12] Tom Leighton. Improving performance on the internet. *Communications of the ACM*, 52:44–51, February 2009.

[13] David Liben-Nowell, Jasmine Novak, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. Geographic routing in social networks. *PNAS*, 102(33):11623–11628, 2005.

[14] Josep M. Pujol, Vijay Erramilli, Georgos Siganos, Xiaoyuan Yang, Nikos Laoutaris, Parminder Chhabra, and Pablo Rodriguez. The little engine(s) that could: scaling online social networks. In *Proceedings of SIGCOMM'10*, pages 375–386, New York, NY, USA, 2010. ACM.

[15] Tiago Rodrigues, Fabrício Benevenuto, Meeyoung Cha, Krishna Gummadi, and Virgílio Almeida. On word-of-mouth based discovery of the web. In *Proceedings of IMC'11*, pages 381–396, New York, NY, USA, 2011. ACM.

[16] Salvatore Scellato, Cecilia Mascolo, Mirco Musolesi, and Jon Crowcroft. Track globally, deliver locally: improving content delivery networks by tracking geographic social cascades. In *Proceedings of WWW'11*, pages 457–466, New York, NY, USA, 2011. ACM.

[17] Salvatore Scellato, Anastasios Noulas, Renaud Lambiotte, and Cecilia Mascolo. Socio-spatial Properties of Online Location-based Social Networks. In *In Proceedings of ICWSM'11*, July 2011.

[18] Sipat Triukose, Zhihua Wen, and Michael Rabinovich. Measuring a commercial content delivery network. In *Proceedings of WWW'11*, pages 467–476, New York, NY, USA, 2011. ACM.

[19] Mike P. Wittie, Veljko Pejovic, Lara Deek, Kevin C. Almeroth, and Ben Y. Zhao. Exploiting locality of interest in online social networks. In *Proceedings of CONEXT'10*, pages 25:1–25:12, New York, NY, USA, 2010. ACM.

[20] YouTube. *Statistics* `http://www.youtube.com/t/press_statistics` Accessed Feb. 2012.

[21] YuMe. Online Video and Television Viewing Attitudes. `http://www.yume.com/sites/default/files/YuMe_Online_Video_Attitudes_Whitepaper.pdf`. Accessed Sep. 2011.

[22] M. Zink, K. Suh, Y. Gu, and J. Kurose. Watch global, cache local: YouTube network traffic at a campus network: measurements and implications. In *Proceedings of MMCN'08*, January 2008.