The Role of Twitter in YouTube Videos Diffusion

George Christodoulou, Chryssis Georgiou, George Pallis

Department of Computer Science
University of Cyprus
{gchris04, chryssis, gpallis}@cs.ucy.ac.cy

Abstract. Understanding the effects of social cascading on streaming media is of great importance to Web information system engineering. Given the large amount of available videos, it is often difficult for users to discover interesting content. Relying on the suggestions coming from friends seems to be a popular way to choose what to watch. Taking into account the increasing popularity of Online Social Networks and the growing popularity of streaming media, in this paper we present a detailed analysis of social cascading exchange of YouTube videos among Twitter users. Using a real data set we have recently collected, our analysis highlights several important aspects of social cascading, including its impact on YouTube videos popularity, dependence on users with a large number of followers, the effect of multiple sharing follows and the distribution of cascade duration.

Keywords: Social Video Sharing, Social Web, Social Cascading, YouTube, Twitter, Internet Measurements

1 Introduction

The rapid proliferation of online social networking sites like Facebook and Twitter has made a profound impact on the Internet and tends to reshape its structure, design, and utility [10]. Industry experts [15] believe that Online Social Networks (OSNs) create a potentially transformational change in consumer behavior and will bring a far-reaching influence on traditional industries of content, media, and communications.

Motivation. Two recent trends in Web information system engineering motivate this work: the increasing popularity of OSNs and the growing popularity of streaming media. In contrast to traditional methods of content discovery such as browsing or searching, OSN sites have recently emerged as a popular way of discovering information on the Web through information dissemination along user's social links. Today OSN sites have been noted as being the primary causes behind the recent increases in HTTP traffic observed in measurement studies. According to Hitwise[1], 8.6% of traffic to news sites now comes from Facebook,

[1] http://www.experian.com/hitwise/index.html

Twitter and smaller social media sites, which is a 57% percent increase since 2009. The percentage coming from search engines, at the same time, is declining; according to Hitwise, it has been observed a drop of 9% since 2009. Social media, in other words, now bring in almost half as much traffic to news sites as search does. A second emerging trend is the growing popularity of streaming media services. The amount of Internet traffic generated every day by online multimedia streaming providers, such as YouTube, is extremely high [18]. Although it is difficult to estimate the proportion of traffic generated by social cascading, it is observed that there are more than 400 tweets per minute with a YouTube link [6].

While the real numbers are debatable, it is clear that the evolution of OSNs and streaming media play a crucial role on Internet traffic, since social cascades (information (i.e., text, image, video) dissemination along links in a OSN [8]) affect the users' navigation behavior. In a recent study [4], authors measured the role that social cascading impacts the diffusion of information. Their experiments showed that social cascades affect significantly the browsing of users. At the same time, YouTube is the most popular and bandwidth intensive service of today's Internet [6]. The mix of the two phenomena has serious implications and presents new challenges for Internet services and content providers towards improving the effectiveness of several services, including caching, content delivery networks, searching and content recommendation [21].

Contributions. In this work we address the following question: *What is the role of social cascading in YouTube video diffusion?* In order to answer this question, we study a large corpus of YouTube videos. For capturing the social cascading effects, we use Twitter, which is one of the most popular OSNs and its core functionality, tweeting, is centered around the idea of spreading information by word-of-mouth [19]. Specifically, Twitter provides mechanisms like retweet (act of forwarding other people's tweets), which enable users to propagate information across multiple hops in the network through cascading. According to a recent announcement[2], Twitter is sharing more than 340 million tweets per day, where 25% of tweets contain links. Overall, this paper makes the following contributions:

– We present the methodology that we have followed in order to collect the Twitter dataset. Our study is based on a newly real dataset from Twitter containing geographic location, follower lists and tweets for 37 million users. Then, we tracked the spreading of more than one million of YouTube videos over this network, analyzing a corpus with more than 2 billions messages and extracting about 1,3 millions single messages with a video link.
– We examine the role of social cascading in YouTube video diffusion. Our analysis highlights several important aspects of social cascading, including its impact on YouTube videos popularity, dependence on users with a large number of followers, the effect of multiple sharing follows and the distribution of cascade duration. Our analysis provides valuable results so as to better

[2] http://mashable.com/2012/03/21/twitter-has-140-million-users/

understand how the retweeting mechanism affects the spread of YouTube videos. To this respect, we introduce a new metric, called *video retweet likelihood*, that measures the likelihood of a user retweeting a video. Although our work has focused only on YouTube videos, its wide popularity and its massive user base allow us to gain insights on user navigation behavior on other similar media platforms.

Roadmap. The rest of this paper is organized as follows. Section 2 reviews previous related work. Our data collection methodology is described in Section 3, whereas, our main findings are presented in Section 4. Section 5 concludes the paper and discusses directions for future work.

2 Related Work

Many studies have been carried to analyze the users' behaviors in different media services [1, 18, 22]. Early work in this area is focused on the analysis and characterization of streaming services in the Internet [22]. In [22], the authors explored workload characteristics based on logs from internal media servers at Hewlett-Packard. Recently, YouTube has been a popular research topic in the Internet measurement community [6, 11]. Several studies [1, 7, 12–14, 17] have been conducted to investigate the traffic characteristics of YouTube users. These works are focused on the characteristics of YouTube content, such as file size, bit-rate, usage patterns and popularity. After an extensive analysis of the YouTube workload in [13], authors found that there are many similarities to traditional Web and media streaming workloads. From another perspective, the authors in [9] studied YouTube videos and found that the videos have strong correlations with each other since the links to related videos generated by uploaders have small-world characteristics. In [11], it is characterized the growth patterns of video popularity on YouTube and analyzed how the popularity of individual videos evolve since the video's upload time.

OSNs are focused on sharing information and as such, have been studied extensively in the context of information diffusion. For instance, the authors in [19] found how propagation of YouTube videos on Twitter is spread among users who are geographically close together. In the same context, the authors in [21] studied how geographic information extracted from social cascades of Twitter can be exploited to improve caching of multimedia files in a Content Delivery Network (CDN). Similarly, in [23], the authors developed a system that exploits information available from Twitter and regularity of activity patterns so as to distribute long-tailed content while decreasing bandwidth costs. In [6], it is measured the popularity distribution of YouTube videos across different geographic regions and analyze how social sharing affects their spatial popularity. According to this study, it is observed that the impact of social sharing on the geographic properties of YouTube video views is significant. Our findings confirm these results with respect to our investigation of the impact of social cascading regarding geographic popularity.

Various studies in the context of social networks have been conducted to predict properties of the social cascading process. The authors in [3] exploited the social cascades in order to identify influencers in Twitter. The authors in [20] focus on characterizing and modeling the information cascades formed by the individual URL mentions in the Twitter so as to predict which users will predict which URL. In a recent study [4], the authors subject 250 million Facebook users to a controlled experiment in order to measure the role that social cascading impacts the diffusion of information.

The present work builds on these earlier contributions in the following key issues. First, whereas the focus of previous studies [1, 12, 14, 17, 18, 22] has been on the analysis and characterization of streaming services in the Internet, we are interested in the analysis of YouTube videos, taking into account the word-of-mouth diffusion of Twitter. Second, although previous works [4, 19] study the impact of social cascading for YouTube videos, they do not focus on a Twitter data set. As we mentioned above, the core functionality of Twitter is centered around the idea of social cascading.

3 Methodology

In this section we first present the methodology we followed in collecting our Twitter data set and then we extract data characteristics from the obtained set. We note that the data collection was not a straightforward process, mainly due to Twitter's recently modified policy of limiting the number of search requests per hour from a given IP address. Below we explain how we managed to collect our data set while respecting, in our opinion, this policy.

3.1 Data Collection

Data collection took over five months using four Cloud infrastructures (Nephelae[3], Okeanos[4], Amazon EC2 and Rackspace). The data are stored locally in a database. Specifically, a Twitter user keeps a brief profile about each user. The public profile includes the full name, the location, a web page, a short biography, and the number of tweets of the user. The people who follow the user and those that the user follows are also listed. In order to collect user profiles, we search for HTTP URLs that were posted on Twitter.

Totally, the data comprises profiles of 37,343,273 users, 6,820,494,777 directed follower links among these users. Due to computation limitations, we select uniformly at random from the above data set 1,384,758 users (247,399,334 directed follower links among these users), whom we focus on in the remainder of this paper. Specifically, 299,071,571 public tweets were posted by these users. The tweets are from December 2011 until April 2012. The period of study allows us to avoid any seasonal side effects exhibited by users navigation behavior. The

[3] Nephelae. http://grid.ucy.ac.cy/Nephelae/
[4] Okeanos. https://cms.okeanos.grnet.gr/about/

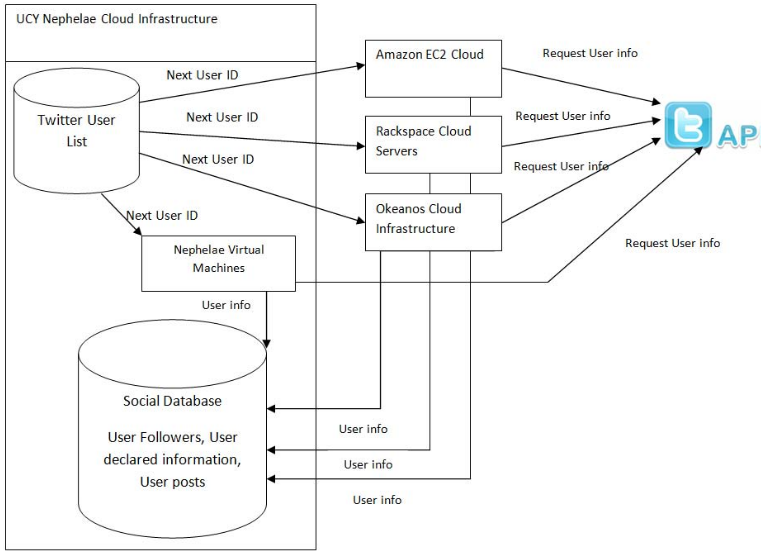

Fig. 1. Data Collection System for Twitter

data set does not include any tweet information about a user who had set his account private. The large-scale of our dataset captures the geographic location diversity of Twitter users. Figure 1 presents the data collection system that we developed in order to collect the Twitter data. In the following paragraphs, we present the methodology that has been followed.

Using the Twitter API[5], we collected user profiles and tweets for each user and then analyze tweets to get HTTP URLs. Twitter imposes rate limiting in the number of search requests per hour from a given IP address. Specifically, one request can fetch up to 5,000 followers for a user, or 200 tweets for a user, or the information of 200 users. To respect this policy, we used several Twitter accounts. Additionally, in order to increase the number of users we used the Twitter social graph obtained in 2009 [16] and requested for Twitter users that were in the Twitter user pool. During our collection period we managed to collect more than 300 million tweets containing HTTP URLs. For each tweet we have crawled the author, the time when it was sent and the actual content of the message. In order to capture the geographic location of users, we used the Geocoding API of Google Maps. The Google Geocoding API provides a direct way to access a geocoder via an HTTP request. Tables 1 and 2 present the information that we collected for each user profile and each tweet respectively.

The next step was to pre-process the content in tweets. While analyzing the URLs within tweets we have found that the majority of the Web links are from URL shortening services (e.g., bit.ly), which substantially shorten the length of any URL. Thus, we used URL shortening services, such as unshort.me, in order

[5] Twitter API. https://dev.twitter.com/

Verified	user's identity has been verified by email account
followers_count	the number of users that follow the user
Protected	user's account is private and only their approved followers can read their tweets or see extended information about them
listed_count	the number of lists the user is a member of
friends_count	the number of users the user follows
Location	the location of user
geo_enabled	if enabled allows applications to send tweets with a geographic location attached
Lang	the language of user
favourites_count	the number of tweets the user has classified as favorites
created_at	the date that the account has been created
time_zone	the time zone of each tweet

Table 1. Twitter User Profile Information

ID	the unique ID of the tweet
Text	the text of tweet (typically up to 140 characters)
created_at	the date that the tweet has been published
Retweeted	if it is new tweet or a retweet
in_reply_to_status_id	the ID of an existing status that the update is in reply to
in_reply_to_user_id	the user ID that the tweet replies
urls	the url of the tweet
retweet_count	the number of times that a tweet has been retweeted

Table 2. Tweet Information

to unshort the URLs. Then, the final step was to gather all the URLs and filter out all the URLs except the YouTube ones.

3.2 Data Set Characteristics

Unlike other OSNs, a Twitter user may follow another user to receive his/her tweets, forming a social network of interest. Furthermore, it is not necessarily the case that two users are mutual followers. Thus, Twitter is represented by a directed graph, where nodes represent the users and a direct link is placed from a user to another user, if the first follows the tweets of the latter. Figure 2 depicts an example of a Twitter social graph. Users A and G are mutual followers, while users A and B are not (A follows B but not vice-versa). According to [19] the node in-degree and out-degree distributions measured on this network are heavy-tailed, and the network topology is similar to those of other OSNs like Facebook. Although a very small fraction of users have an extremely large number of friends, the majority of users have only a few friends. The most

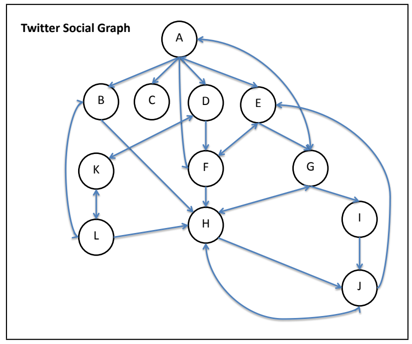

Fig. 2. An example of the representation of Twitter followers as a social graph.

popular users act as authorities and are usually either public figures or media sources. These observations are confirmed by the data set we have collected. Another interesting observation from our Twitter data set is that the high in-degree nodes are not necessarily high out-degree nodes since a small number of links are bidirectional (that is, followers are mutual).

4 Effects of Social Cascading

In this section we investigate the role of the retweeting mechanism in YouTube video diffusion. To this respect, social cascades include only users that have tweeted a certain video link. Also, without loss of generality, we make the assumption that every video contained in a Twitter message has been viewed by the user who retweet it (in other words, we assume that users do not "blindly" retweet videos).

To measure the impact of social cascading, and more specifically of video retweeting, we define the *video retweet likelihood* metric over our data set, as the likelihood of retweeting a YouTube video. More formally, we define the set $OutgoingVideos$ to be the set of videos (over all users) that were *retweeted* by some user to another user (multiplicities are not counted). Similarly, we define the set $IncomingVideos$ to include all videos that a user received by some other user, as a result of a *retweet* of the latter. Then, the *video retweet likelihood* is computed by the following expression:

$$\frac{|OutgoingVideos \cap IncomingVideos|}{|IncomingVideos|}.$$

Note that this metric captures only the videos that have been retweeted. The cardinality of the intersection gives the number of videos that users retweeted among the ones that were retweeted to them. Dividing this number with the total number of "received retweeted" videos gives the likelihood of a video being retweeted.

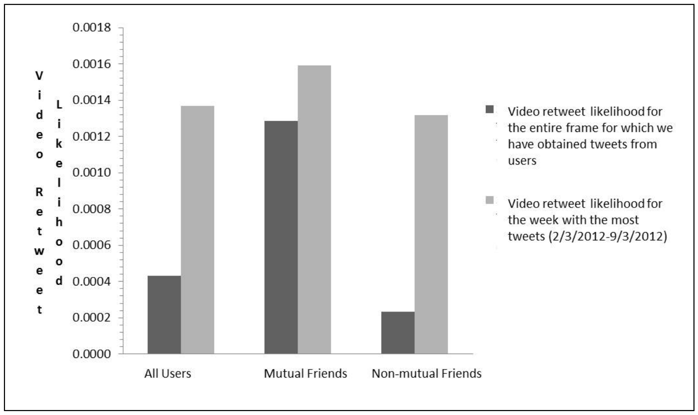

Fig. 3. Influence of *video retweets* with respect to the relationship between Twitter Users

4.1 Impact of Twitter Users

First, we study how retweeting influences the diffusion of YouTube videos. Our aim is to investigate the impact of social cascading on the users' navigation behavior. In general, it is more likely to view and retweet a video that has been tweeted by a follower. Considering that each tweet can be viewed by all the followers of the author, the potential audience that a YouTube video may reach via retweeting is much larger, even if only few users are involved. As displayed in Figure 3, an interesting observation is that the *video retweet likelihood* is 6 times larger for users that are mutual followers. Recall that two users are mutual followers if the relationship of following and being followed is reciprocal (like users A and G in Figure 2). Moreover, we studied the *video retweet likelihood* for the week with the most tweets. Comparing with the results that we took for the whole period, to our surprise, we observe a different behavior for the non-mutual followers. On the other hand, the results for mutual followers are quite similar for both periods.

Furthermore, in Figure 4 we show how the *video retweet likelihood* is affected taking into account the number of users that have shared a tweet. We observe that the *video retweet likelihood* is increased with the number of users' follows who have already shared the same tweet. This increase seems to be exponential when the same tweet is shared by more than 8 follows. This is consistent with recent observational studies in other OSNs, such as Facebook [4].

4.2 Impact of Geographic Popularity

Our next study is to investigate the impact of social cascading regarding the geographic popularity. To capture the geographic popularity, we use the *time-zones* of users. Twitter enables users to declare their time zone. However, instead of using the UTC time zone system (where the globe is divided into 24 time zones),

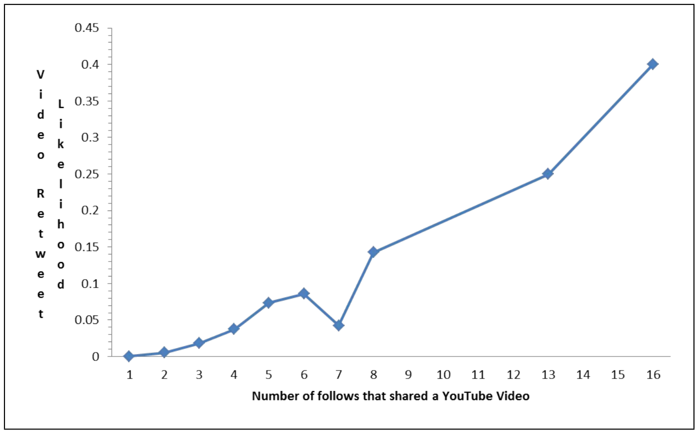

Fig. 4. The *video retweet likelihood* with respect to the number of follows that share a YouTube video.

Twitter uses its own time zone system which divides the globe into 142 zones. In our study we consider that the users with the same Twitter time-zone have the same geographic location. Then, according to our data set, the average number of users per zone is 93,076.3, the median 7,167 and the standard deviation is 369,823.6 (this large number is due to the different population distribution of different time zones: some zones expand over entire countries, where others over small cities). Figure 5 depicts the *video retweet likelihood* with respect to the population in logarithmic scale. Each point in the plot depicts a group of users that belong in the same time zone, since users are mainly influenced by follows who are in the same geographic location. Our results show that the smaller the population is, the larger the *video retweet likelihood* is. This means that the social cascading effect has high impact on a more focused and less diverse set of geographic regions. These findings confirm recent results [2, 6] which found that geographic distance affects social interaction on OSNs. Specifically, the highest *video retweet likelihood* (0.1) has been observed in the region of Astana. Astana is a place with 700,000 habitants, whereas, the average likelihood is 0,0016.

Furthermore, we study the relationship between the popularity of videos with respect to the number of retweets. In order to understand the impact of social cascading for videos with higher number of views we have classified all videos regarding their popularity (this statistic is given by the YouTube API) into 16 segments (of views). The results (Table 3) show that the popularity versus the number of retweets is not trivial. The effect of social cascading for these groups of videos is different across these 16 categories. The general trend is that the more popular a video is, the more retweets has. From Table 3, it occurs a surge in the number of retweets for very popular videos. Also, it is interesting that the average number of retweets does not exceed an upper limit, which is 26 cascades in our case.

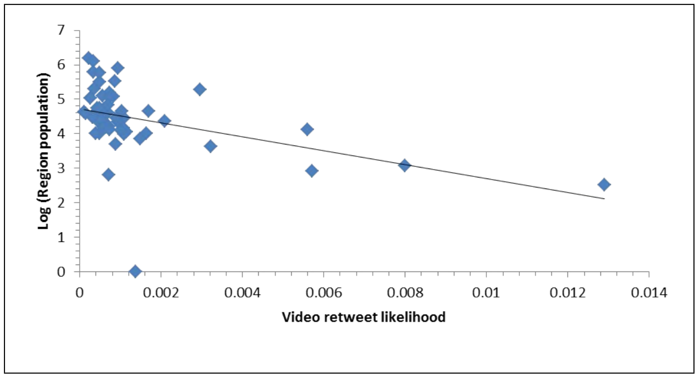

Fig. 5. The log plot geographic location population of YouTube videos with respect to the *video retweet likelihood*.

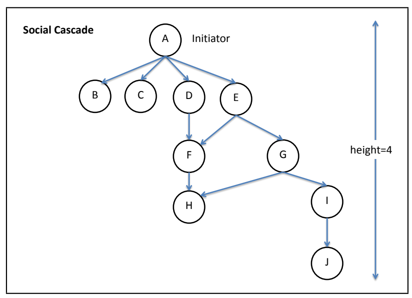

Fig. 6. An example of a cascade represented as a tree rooted at the initiator.

4.3 Impact of Social Cascade Length

The next experiment is to study the *video retweet likelihood* with respect to the length of the cascade. Cascades can be represented as rooted directed trees where the initiator of the cascade is the root of the tree [3]. Figure 6 depicts an example of a cascade, initiated by user A over the Social Graph of Figure 2. Then, the *length* of the cascade is the *height* of the resulting tree (which is 4 in Figure 6).

Figure 7 depicts the distribution of the cascade length (given in log scale), which is approximately power-law. This measure of popularity demonstrates that it is rare to have large cascades, but when they do take place they can become extremely large. This implies that the vast majority of posted YouTube videos do not spread at all.

4.4 Impact of Time

In Figure 8, we illustrate the distribution of cascade duration (in hours) from the first tweet to the last tweet for each cascade with at least 2 users, not counting the initiator. This result shows how YouTube links can spread on Twitter on a

Views	Number of videos	Number of retweets	Avg. Number of retweets
1000	68655	85252	1.24
5000	37899	45640	1.20
20000	43014	49197	1.14
50000	34855	40515	1.16
200000	53509	68099	1.27
400000	23544	34638	1.47
700000	16010	27011	1.68
1 million	8571	16452	1.92
2 millions	13332	27721	2.08
5 millions	11183	30245	2.70
10 millions	4641	18924	4.07
20 millions	2205	12967	5.88
50 millions	1210	10775	8.90
100 millions	393	10519	26.76
200 millions	21	553	26.33
350 millions	7	176	25.14

Table 3. Popularity of YouTube videos

time scale. About 70% of the cascades end within 24 hours. In particular, about 25% of the cascades occur within the first hour, in 3 hours the spread reaches to 40% and about 85% of the cascades end by the third day (72 hours). This indicates that links to videos can quickly spread over the social network, leading to many views in a short period of time. This information could be exploited, for example, in improving the efficiency of Content Delivery Networks, as discussed in the next section.

5 Conclusion and Future Work

The widespread adoption of OSN sites has significantly altered the information diffusion through the Web. In this work, we have presented how the retweeting influences the diffusion of YouTube videos. Using an experimental approach on Twitter, we are able to quantify the effect of social cascading on video spread. This study is useful for Internet service and content providers, who can exploit these findings towards improving the effectiveness of their services.

One of the most sound observations of our study is that the social cascading effect has high impact on a more focused and less diverse set of geographic regions. Also, the social cascading effect ends within 24 hours. These findings are useful for large-scale systems whose traffic is driven by online social services. For instance, Content Delivery Networks (CDNs) can take advantage of the fact that social cascades can spread in a geographically limited area to decide whether a YouTube video is disseminating locally or globally.

Another interesting observation is that social cascading affects the users navigation behavior. In the case of Twitter, users are influenced more from the follows

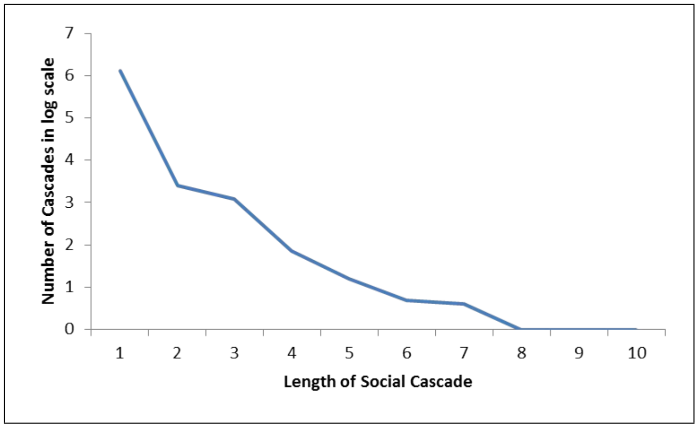

Fig. 7. Number of social cascades in log scale with respect to the length of the social cascade

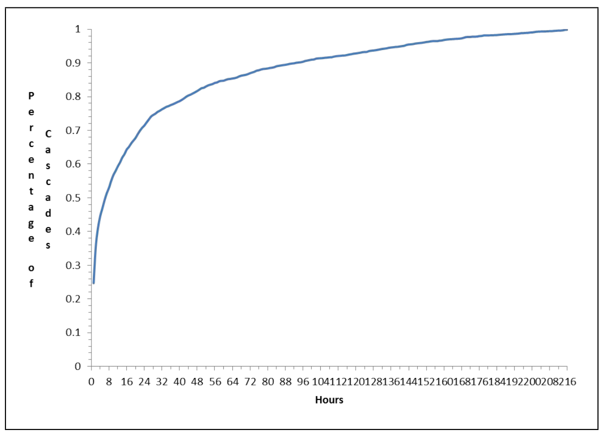

Fig. 8. The cumulative distribution function of the social cascades with respect to the time passed.

who also follow them as well. This finding may be used to study the role of influencers in Twitter. A related work in this area has been presented in [3]. Also, our analysis showed that most events through Twitter do not spread at all, and even moderately lengthed cascades are extremely rare.

For the future, we plan to further investigate the impact of social cascading in YouTube video diffusion. Specifically, we will study the retweeting influence with respect to the popularity dynamics of YouTube videos over the time [5]. An implication of this study is the improvement of Internet-based content delivery. The rapid proliferation of OSNs opens new perspectives in Internet-based content technologies, raising new issues in the architecture, design and implementation of existing CDNs. In this context, we plan to develop a realistic media workload generator that would reflect the dynamics and evolution of content at media sites and the change of access rate to this content due to the role of social networks in information diffusion. The media workload generator will produce synthetic traces with desired distributions and controllable parameters for performance experiments studying effective streaming content delivery approaches. The ultimate goal of this generator is to be used as a valuable tool in order to study efficient algorithms towards predicting social cascades and improving the performance of CDNs.

Acknowledgement: The authors thank the anonymous reviewers for their constructive comments and suggestions. This work is supported in part by the Startup Grant [George Pallis, 2011], funded by the University of Cyprus.

References

1. A. Abhari and M. Soraya. Workload generation for youtube. *Multimedia Tools Appl.*, 46(1):91–118, Jan. 2010.
2. L. Backstrom, E. Sun, and C. Marlow. Find me if you can: improving geographical prediction with social and spatial proximity. In *Proceedings of the 19th International Conference on World Wide Web (WWW '10)*, pages 61–70, New York, NY, USA, 2010. ACM.
3. E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts. Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the 4th ACM International Conference on Web Search and Data Mining (WSDM '11)*, pages 65–74, New York, NY, USA, 2011. ACM.
4. E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic. The role of social networks in information diffusion. In *Proceedings of the 21st International Conference on World Wide Web (WWW '12)*, pages 519–528, New York, NY, USA, 2012. ACM.
5. Y. Borghol, S. Mitra, S. Ardon, N. Carlsson, D. Eager, and A. Mahanti. Characterizing and modelling popularity of user-generated videos. *Perform. Eval.*, 68(11):1037–1055, November, 2011.
6. A. Brodersen, S. Scellato, and M. Wattenhofer. Youtube around the world: geographic popularity of videos. In *Proceedings of the 21st International Conference on World Wide Web (WWW '12)*, pages 241–250, New York, NY, USA, 2012. ACM.
7. M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. Moon. I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement (IMC '07)*, pages 1–14, San Diego, California, USA, 2007. ACM.
8. M. Cha, A. Mislove, B. Adams, and K. Gummadi. Characterizing social cascades in flickr. In *Proceedings of the first workshop on Online social networks (WOSN '08)*, pages 13–18, Seattle, WA, USA.
9. X. Cheng, C. Dale, and J. Liu. In *Proceedings of the 16th International Workshop on Quality of Service (IWQOS '08)*, pages 229–238, Enskede, The Netherlands, 2008. IEEE.
10. A. Datta, M. D. Dikaiakos, S. Haridi, and L. Iftode. Infrastructures for online social networking services. *IEEE Internet Computing*, 16:10–12, 2012.
11. F. Figueiredo, F. Benevenuto, and J. M. Almeida. The tube over time: characterizing popularity growth of youtube videos. In *Proceedings of the 4th ACM*

International Conference on Web Search and Data Mining (WSDM '11), pages 745–754, New York, NY, USA, 2011. ACM.
12. A. Finamore, M. Mellia, M. M. Munafò, R. Torres, and S. G. Rao. Youtube everywhere: impact of device and infrastructure synergies on user experience. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference (IMC '11)*, pages 345–360, New York, NY, USA, 2011. ACM.
13. P. Gill, M. Arlitt, Z. Li, and A. Mahanti. Youtube traffic characterization: a view from the edge. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement (IMC '07)*, pages 15–28, San Diego, California, USA, 2007. ACM.
14. P. Gill, M. Arlitt, Z. Li, and A. Mahanti. Characterizing user sessions on youtube. In *Proceedings of the ACM/SPIE Multimedia Computing and Networking Conference (MMCN '08)*, San Jose, USA, 2008.
15. S. Goel, J. M. Hofman, and M. I. Sirer. Who does what on the web: A large-scale study of browsing behavior. In *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media*, June 2012.
16. H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a social network or a news media? In *Proceedings of the 19th International Conference on World Wide Web (WWW '11)*, pages 591–600, New York, NY, USA, 2010. ACM.
17. S. Mitra, M. Agrawal, A. Yadav, N. Carlsson, D. Eager, and A. Mahanti. Characterizing web-based video sharing workloads. *ACM Trans. Web*, 5(2):8:1–8:27, May 2011. ACM.
18. A. Rao, A. Legout, Y.-s. Lim, D. Towsley, C. Barakat, and W. Dabbous. Network characteristics of video streaming traffic. In *Proceedings of the 7th Conference on Emerging Networking Experiments and Technologies (CoNEXT '11)*, pages 25:1–25:12, New York, NY, USA, 2011. ACM.
19. T. Rodrigues, F. Benevenuto, M. Cha, K. Gummadi, and V. Almeida. On word-of-mouth based discovery of the web. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference (IMC '11)*, pages 381–396, New York, NY, USA, 2011. ACM.
20. D. M. Romero, W. Galuba, S. Asur, and B. A. Huberman. Influence and passivity in social media. In *Proceedings of the 20th International Conference on World Wide Web (WWW '11)*, pages 113–114, New York, NY, USA, 2011. ACM.
21. 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 the 20th International Conference on World Wide Web (WWW '11)*, pages 457–466, New York, NY, USA, 2011. ACM.
22. W. Tang, Y. Fu, L. Cherkasova, and A. Vahdat. Medisyn: a synthetic streaming media service workload generator. In *Proceedings of the 13th International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV '03)*, pages 12–21, New York, NY, USA, 2003. ACM.
23. S. Traverso, K. Huguenin, I. Triestan, V. Erramilli, N. Laoutaris, and K. Papagiannaki. Tailgate: handling long-tail content with a little help from friends. In *Proceedings of the 21st International Conference on World Wide Web (WWW '12)*, pages 151–160, New York, NY, USA, 2012. ACM.