

# The Propagation and Influence of A Web Video using Unified Virtual Community Space

B. Saranya\*<sup>1</sup>

, P.G. Scholar,

Department of Computer Science & Engineering,  
MNSK College of Engineering,  
Pudukkottai – 622305, Tamilnadu, India

G. Sathishkumar\*<sup>2</sup>

Assistant Professor, HOD

Department of Electronics & Communication Engineering  
MNSK College of Engineering,  
Pudukkottai – 622305, Tamilnadu, India

**Abstract**—Our project signifies propagation analysis on an open social network, i.e., YouTube, by crawling one of its friendship networks and one of its subscriber's networks. Propagation of web videos that is by estimating two things. One is the estimation of video influences i.e. how the video plays a role in everyone's life, it may be a public issue or a private issue. Second is the estimation of origin of videos i.e. how the video get starting origination into outside world. It may be get originated from a popular site or from some particular person's website and get popularity. We observed that the effect on propagation of people who are not either in a friendship network or a subscription network is higher than that of friends or subscribers. Unified Virtual Community Space is proposed to model the propagation of the video. To calculate the video origin and influence Noise-reductive Local-and-Global Learning method is used and in the experiments to test the video from different sites and find the propagation of the videos.

**Key words:** *Unified Virtual Community Space, Web video, Influence*

## 1. INTRODUCTION

We propose a novel approach to analyze how a popular video is propagated in the cyberspace, to identify if it originated from a certain sharing-site, and to identify how it reached the current popularity in its propagation. In addition, this project also estimates their influences across different websites outside the major hosting website. Web video is gaining significance due to its rich and eye-ball grabbing content. When a video receives some degree of popularity, it tends to appear on various websites including not only video-sharing websites but also news websites, social networks or even Wikipedia. Numerous video-sharing websites have hosted videos that reached a phenomenal level of visibility and popularity in the entire cyberspace. As a result, it is becoming more difficult to determine how the propagation took place - was the video a piece of original work that was intentionally uploaded to its major hosting site by the authors, or did the video originate from some small site then reached the sharing site after already getting a good level of popularity, or did it originate from other places in the cyberspace but the sharing site made it popular.

### 1.1 WEB VIDEOS

As the Web continues to evolve, one of the most noticeable phenomenon is the prevailing of videos as a major source of multimedia information on the Web. The latest research conducted by comScore1 reveals that in the single

month of January 2011, a U.S. Internet user spent 870.8 minutes in average on viewing web videos. Web videos nowadays influence society likes never before in history. As we witnessed the success of YouTube, Hulu and other video-sharing websites, we have also noticed how the social networks have fueled the growth of online videos. When a video becomes popular, it can be spotted not only on one or more video-sharing websites but also on news websites, social networks, blogs or even Wikipedia.

Vice versa, when a video from news websites, social networks or blogs gets attention, it is likely to be put on video-sharing sites as well. In such context, it is utterly important to identify how the propagation took place, i.e., to determine if a popular video on a video sharing website actually originated from that website, or it is merely a projection of influence from somewhere else of the cyberspace. Particularly, in this study we primarily focus on the identification of the propagation patterns of the web videos. We also study their influence in the entire cyberspace.

The problem we aim to solve is partially similar to the analysis a user's friends and the identification of his/her influence in a social network, but there are some key differences. In influence analysis in social network, all users, or nodes from a network's perspective, are normally considered to be in a single website, in which a user's influence can be identified with existing approaches by analyzing the friend relationships and interactions with other users. In such case, the concept of origin for a user does not exist.

However, the problem becomes more difficult if we consider an online video's propagation and influence as in this case multiple websites need to be examined. On the one hand, a video's existence on a hosting site may be affected by some emerging events from other websites. On the other hand, a video originating from a hosting site makes its way to the most popular video inside the site, and then draws dramatic attention from other websites.

Fig. 1.1 shows the most viewed ten videos in all time from the largest online video-sharing site YouTube.com. After close investigation of the videos' propagation in cyberspace, we conclude that video 1, 9 and 10, which are marked with stars, are the origins of other duplicate videos in the cyberspace.

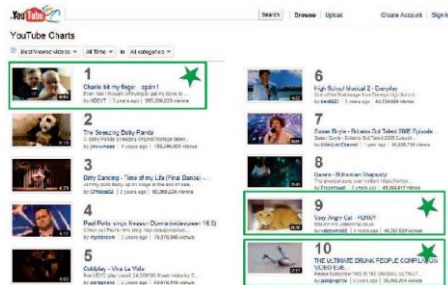


Fig.1.1: Popular web videos.

These videos originate from YouTube.com, and are then propagated to the cyberspace via other websites. During the process, they have drawn remarkable public attention from both YouTube.com and other websites. A very good example is the video Charlie Bit My Finger - again2. As the No. 1 web video with the most views in history (up to 2010), the video is widely received and reported, on Wikipedia, MySpace, Twitter, personally blogs, and numerous news websites like The Telegraph<sup>3</sup>, Time<sup>4</sup> and Sydney Morning Herald<sup>5</sup>. Clearly the video was firstly uploaded on YouTube.com and then became publicly popular on other websites.

In this case, its original hosting site outputs great influence to the cyberspace rather than receiving influence from it. That means, the propagation of this video on other websites demonstrates a video sharing site's significance as an information source. We illustrate this propagation in Fig.1.2. Contrarily, some other videos show a different case with respect to the propagation of influence. For instance, Coldplay - Viva La Vida, as the No. 5 video, is popular in the YouTube U.K. community.

The video is a duplicate of a music video which is already widely hosted on other websites, thus YouTube is not the very source of the influence. When searching with the video title on major search engines, the dominant part of relevant entries links to the pages concerning the song but not the video itself. Little attention is brought to the video's hosting site compared to the previous example. In such cases, we say the popularity that the video receives on YouTube is a co-effect of the public popularity of the song itself, and YouTube is not the origin of the propagation or influence. The same observation can be obtained for other unmarked videos in Fig.1.1 Our objective is to analyze the propagation as well as the direction of influence for a video, and then evaluate the influence in the public domain.

The problem we target to solve is extremely important to the video sharing site owner, in the following scenarios.

Some video-sharing sites have developed schemes to encourage users to upload content. For instance, YouTube.com decided to give cash rewards to successful video uploaders<sup>6</sup>.

As encouraging as it can be for individual content producers, it also poses challenges and issues when a video is produced somewhere else but is shared on the sites without proper permission. Though this type of videos can also become popular in the sharing sites, such action should not be encouraged or even rewarded for. Our study facilitates such decision making process by providing means to analyze a user's uploaded videos and determine if the user's uploading activities have been legitimate and if the user should be rewarded.

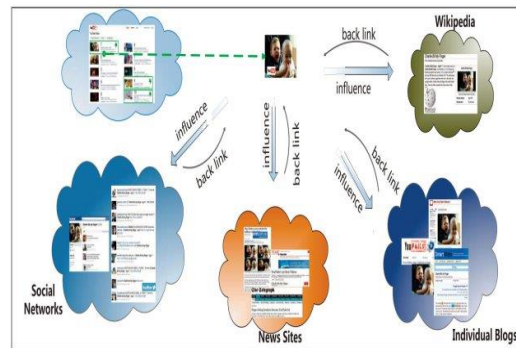


Fig.1.2: Influence propagation of the video

To model an online video's propagation and influence in the cross-community cyberspace, we define a Unified Virtual Community Space that captures the propagation history of an online video. The UVCS records key information of an online video, such as the video page's ranking in the search results for a text query with the video's title on search engines, and the information about the video page's inbound and outbound links, etc. UVCS is used as the raw feature for our algorithm to classify the propagation and rank the influence of an online video. A video's UVCS is independent from another video's UVCS.

We propose an advanced learning method called Noise-reductive Local-and-Global Learning to fulfill the following learning objectives: – The method should be able to reduce noise. The UVCS feature is a combination of multiple semantic components. The significance of each component is not specified in the raw feature. Fields of the UVCS feature may be missing for some feature vectors due to the diversified nature of web pages. Overall the feature is regarded very noisy.

## 1.2 PROBLEM FORMULATION AND FRAMEWORK OVERVIEW

We formulate the problem and describe the general framework. We give the following formulation of the problem studied here.

Formulation: Given a set of videos

$V = \{v_i\}$ , establish corresponding distinctive features  $X = \{x_i\}$  to describe their patterns of propagation and overall influence, classify their propagation patterns into  $C = \{c_i\}$  and evaluate their influence scores  $S = \{s_i\}$ , which are interpreted by the criteria presented in Table 1.3.

Fig.1.3 outlines the proposed framework. The framework starts with the most popular videos retrieved from a particular sharing site that we aim to analyze. Those videos receive the most attention, which is reflected by their view counts, ratings and discussions, so they are collected as candidates. For each video candidate, its title text is used as search terms to search on search Engines

The reason why we used this technique is threefold: firstly it involves dramatic effort to crawl all possible web pages and identify duplicate imbedded videos to identify relevant pages; secondly some of the pages only have text reference to the video but have no links or actual embedded

video on them; finally in our investigation we find that most of the relevant pages, with or without video links, could be accessed through text search. Missing a video's true origin in the text search engine results is highly unlikely.

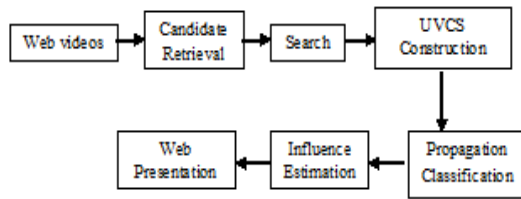


Fig.1.3: Framework.

Hence we argue that this technique is effective enough for our task. The pages returned by the search engines are analyzed, based on which a corresponding feature vector in the Unified Virtual Community Space (UVCS) is constructed. The UVCS is a feature space that consists of elements relevant to a video's propagation and influence, including the link relations of relevant pages, and the tracking of its presence on other websites, e.g. Twitter, Wikipedia, the Blogosphere, and the news websites. We formulate the feature for the candidate video so that the features are used in the NLGL algorithm.

Score range	Criteria
[0,0.25]	The video can be found on some social networks, individual websites but not widely available. It cannot be found on influential media outlets.
[0.25,0.5]	The video is widely available on major social networks, numerous individual websites cannot be found on influential media outlets.
[0.5,0.75]	The video is widely available on major social networks, numerous individual websites, how can only be found on either a news website or Wikipedia
[0.75,1]	The video is widely available on major social networks, numerous individual websites, and can be found on multiple influential media outlets, like news websites, Wikipedia, etc

Table 1.3: Public Influence Scores

### 1.3 UNIFIED VIRTUAL COMMUNITY SPACE

In a social network, the user influence is often modeled as a network flow problem, or similarly, as a link analysis problem. The outbound links and inbound links for a user, established by comments and friend connections, are considered to assess the user's role and influence within the social network. The existing approaches focus on the estimation of within-community influence of an individual user or a group. Its analysis benefits from three factors for being from a single and unified community. These include the unified data format, the same scale and meaning of indicators, and the ease of constructing the link graph to monitor the

propagation. However, in our case, which is a cross-community analysis, these helpful factors are not valid any more.

For example, for the video-sharing websites we are interested in the video's statistics like view count, uploaded time, location, etc. While for news websites or Wikipedia, we only need to find out the reference activities to the video. They also show different scales in data and user numbers, in addition their link relations among the web pages from different communities are more difficult to discover than in the single community case. Given above characterization, to reasonably represent the relevant factors of a video's propagation in cyberspace, we define the concept of the UVCS which contains the web pages relevant to the video. In this UVCS, all the pages have relevance to the video, and they have some intrinsic properties:

- Each page has a time stamp, indicating the publishing time or modification time of the page.
- Each page receives a set of inbound links. The link relations among pages inside the UVCS and pages outside of UVCS are ignored. That means we count only the links from the pages in the UVCS to the pages in the UVCS. A graph with pages that are related to the same video is then established.
- Each page's rank in the UVCS is known. Provided by the search engine, this factor mainly reflects its importance of the website that hosts the page. Combined with the previous item, this item describes how important this page is from a PageRank type of view. Nevertheless this factor does not provide much information about the direction of influence.

Other than these properties, complementary information collected from other sources, i.e., related blog posts, tweets, news, encyclopedia pages, can be retrieved using specialized search engines. The properties and the complementary information described above roughly depict the history of a video's propagation and the evidence of its public influence.

Hence time is an important but not a decisive factor. Similarly, a page that ranks the highest in the UVCS is not necessarily the origin of all similar videos in cyberspace, as its superior rank may be the product of its host's great influence in cyberspace instead of its own origin.

The same case applies for other components in the UVCS. Apparently when identifying the propagation and influence of a video, all these components need to be considered simultaneously. Further analysis needs to be performed on this UVCS to determine the propagation and influence for an online video. Next we describe how the UVCS for a video is constructed and represented. Note that each video has its own individual UVCS.

In other word, we describe a video's propagation with a separate UVCS in which it lives. Each video and its correspondent UVCS is independent from other videos and their UVCS. Such characteristics entitles the whole framework another advantage, that any learning model we trained will be able to be applied on out-of-sample videos, given that the new UVCS are constructed for the out-of-sample videos. Next we show how to construct the UVCS for a given video.

## 2. LITERATURE REVIEW

### 2.1 Manifold regularization: A geometric framework for learning from labeled and unlabeled examples

Author's Name: M. Belkin, P. Niyogi, and V. Sindhwani

- Classify the labeled and unlabeled data from the data set.
- Unlabeled data, when used in conjunction with a small amount of labeled data, can produce considerable improvement in learning accuracy.
- Find out the similarities between videos present in the website.

### 2.2 Understanding video interactions in YouTube

Author's Name: F. Benevenuto

- Used to find out the response for the uploaded video.
- Determine the importance of the user based on the response given for the video.
- Web page has a high rank if the page has many incoming links or a page has links coming from highly ranked pages.

### 2.3 I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System

Author's Name: Meeyoung Cha, HaewoonKwak, Pablo Rodriguez, Yong-YeolAhn, and Sue Moon

- Self-publishing the videos on the site.
- Known information about how to avoid the multiple copies of the video.

### 2.4 Efficient algorithms for ranking with SVMs

Author's Name: O. Chapelle and S. S. Keerthi

- Rank calculation of the video
- Web page ranking-To identify the page contains any internal links or not.based on the link the rank is calculated.
- Methods to know, how to increase the score for relevant document when compared to the irrelevant document present in the page.

### 2.5 Support-Vector Networks

Author's Name: C. Cortes and V. Vapnik

- Separation of the video based on the class label.
- Map the video in the cyberspace.
- Learning information about the video.

## 3. PROJECT DESCRIPTION

### 3.1PROJECT OVERVIEW

On web it is important to identify how the propagation took place, i.e., to determine if a popular video on a video sharing website actually originated from that website, or it is merely a projection of influence from somewhere else of the cyberspace. Particularly, in this study we primarily focus on the identification of the propagation patterns of the

web videos. Also study their influence in the entire cyberspace. The problem we aim to solve is partially similar to the analysis a user's friends and the identification of his/her influence in a social network, but there are some key differences. In influence analysis in social network, all users, or nodes from a network's perspective, are normally considered to be in a single website, in which a user's influence can be identified with existing approaches by analyzing the friend relationships and interactions with other users. However, the problem becomes more difficult if we consider an online video's propagation and influence as in this case multiple websites need to be examined. Due to the open nature of the Web, an online video's influence often exhibits a bi-directional fashion. On the one hand, a video's existence on a hosting site may be affected by some emerging events from other websites. On the other hand, a video originating from a hosting site makes its way to the most popular video inside the site, and then draws dramatic attention from other websites.

### 3.2 EXISTING SYSTEM

Existing research usually concerns the propagation or influence of topics and users in a single community. This project surveys the literature in these related fields, and reviews some of the relevant machine learning methods to our proposed NCRC framework. First this project surveys the recent studies in event discovery and other mining tasks in social networks. Twitter is visualized as a sensor network for event detection. With semantic analysis of tweets; it establishes a probabilistic model which derives the probability of an emerging event from the occurrence reading of related tweets. Then with the assistance of Bayesian filters, it determines the location of the event. It is reported effective for earthquakes, typhoons, or even new video game releases. Meanwhile, A statistical model called PET is defined to track events in social networks. It models the events over time, and exploits the bursts of user interest, the network structural information and the evolution of a topic for event tracking. The approach uses the query likelihood and news headline prior for top news identification in the Blogosphere.

#### Disadvantages

- It concerns the propagation or influence of topics and users in a single community.
- The method used here is transductive, which cannot be applied to out-of-sample data.
- Difficult to perform the video relevance ranking and video thread tracking.

### 3.3 PROPOSED SYSTEM

We define a Unified Virtual Community Space that captures the propagation history of an online video. The UVCS records key information of an online video, such as the video page's ranking in the search results for a text query with the video's title on search engines, and the information about the video page's inbound and outbound links, etc. UVCS is used as the raw feature for our algorithm to classify the propagation and rank the influence of an online video. A video's UVCS is independent from another video's UVCS. We propose an advanced learning method called Noise-



reductive Local-and-Global Learning to fulfill the following learning objectives. The UVCS feature is a combination of multiple semantic components; the significance of each component is not specified in the raw feature. Fields of the UVCS feature may be missing for some feature vectors due to the diversified nature of web pages.

Advantages

- It follows inductive learning, which is very efficient so that it is possible to use it to handle large-scale data.
- Our method simultaneously reduces the noise in the data by dimension reduction.
- Easily determine if a popular online video originated from the video sharing site, or from somewhere else of the Internet

#### 4. IMPLEMENTATION

##### 4.1 CANDIDATE RETRIEVAL

In this module, we are going to perform candidate retrieval process. We are retrieving the reasonable content and show various provider link based on the candidate entered keyword. Candidate wants to access this any one link just click the URL and getting the various information from corresponding service provider.

##### 4.2 SEARCH VIDEOS

In this module are used retrieving the various videos formats from different videos sites. Also help of analyzing the video propagation and estimation process for various network sites. Normally web displays various types of content but our module help of only retrieval in videos formats depends on the filtering process.

##### 4.3 UVCS CONSTRUCTION

The Unified Virtual Community Space captures the propagation history of an online video. The UVCS records key information of an online video, such as the video page's ranking in the search results for a text query with the video's title on search engines, and the information about the video page's inbound and outbound links, etc. UVCS is used as the raw feature for our algorithm to classify the propagation and rank the influence of an online video. A video's UVCS is independent from another video's UVCS.

##### 4.4. PROPAGATION CLASSIFICATION

In this module, we are classifying the video influence propagation. Normally video content displays on the various social service networks (Wikipedia, Facebook, and YouTube). NLGL is designed to fit in our application scenario. In our application, we will gather a collection of UVCS features for the popular online videos, however only a small portion of them will be annotated by expert annotators due to limited human resources.

##### 4.5. INFLUENCE ESTIMATION

The search engines are analyzed based on which a corresponding feature vector in the Unified Virtual Community Space is constructed. The UVCS is a feature space that consists of elements relevant to a video's

propagation and influence, including the link relations of relevant pages, and the tracking of its presence on other websites, e.g. Twitter, Wikipedia, Blogosphere and the news websites. We formulate the feature for the candidate video so that the features are used in the NLGL algorithm.

#### 5. CONCLUSION

Online videos are so popular nowadays that they begin to change people's way of daily entertainment greatly. We presented characteristics of video sharing propagation in social networking services and how online videos propagate and how influential they are outside a video sharing site is an increasingly significant research problem. The identification of an online video's origin and propagation patterns, from the video sharing site's perspective, is crucial to its business models as well to its partner's decision making for their marketing strategies. We identified different types of users in video propagation and evaluate their activities. The UVCS utilizes multi-modal indicators include spatial information, page inter-linkage relations, social network and news media exposure, and so on. Subsequently it offers a comprehensive and panoramic way of describing an online video's life cycle. Then we devise a novel learning method called NLGL. NLGL exploits the benefits of local learning, manifold structure and dimension reduction.

#### 6. FUTURE ENHANCEMENT

In our future research, there is great potential for us to extend this research and enable it with other interesting capabilities. The most likely problem would be the identification of the actual origin of the video. We are also investigating possibilities of establishing inter-sharing-site influence model to analyze how video sharing site influence each other.

#### REFERENCES

1. M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 7, pp. 2399–2434, Nov. 2006.
2. F. Benevenuto et al., "Understanding video interactions in youtube," in *Proc. ACM Multimedia*, New York, NY, USA, pp. 761–764, 2008.
3. M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. B. Moon, "I tube, you tube, everybody tubes: Analyzing the world's largest user generated content video system," in *Proc. IMC*, San Diego, CA, USA, pp. 1–14, 2007.
4. O. Chapelle and S. S. Keerthi, "Efficient algorithms for ranking with SVMs," *Inform. Retr.*, vol. 13, no. 3, pp. 201–215, 2010.
5. C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
6. S. Ji, L. Tang, S. Yu, and J. Ye, "Extracting shared subspace for multi-label classification," in *Proc. 14th KDD*, Las Vegas, NV, USA, pp. 381–389, 2008.
7. R. Ji, X. Xie, H. Yao, and W.-Y. Ma, "Mining city landmarks from blogs by graph modeling," in *Proc. ACM Multimedia*, Beijing, China, pp. 105–114, 2009.