B. M. S. COLLEGE OF ENGINEERING

(Autonomous Institute, Affiliated to VTU, Belagavi)

Post Box No.: 1908, Bull Temple Road, Bengaluru – 560 019

DEPARTMENT OF MACHINE LEARNING

Academic Year: 2022-2023 (Session: October 2022 - February 2023)

SOCIAL MEDIA ANALYTICS (22AM6PESMA)

SMA AAT-4

Submitted by

Student Name:	Monesh S, Somansh V
USN:	1BM20AI039, 1BM20AI050
Date:	**07/06/2023**
Semester & Section:	**6 A**
Total Pages:	
Student Signature:	*somansh, monesh*

Valuation Report (to be filled by the faculty)

Score:	
Faculty	**Dr. Soumya Lakshmi**
Faculty Sign	

1. Introduction

In the modern digital landscape, social media platforms have become integral to our lives, providing avenues for communication, information sharing, and community building. With billions of users and an ever-increasing volume of content being shared daily, social media platforms have evolved into valuable sources of data and insights for businesses, researchers, and marketers alike.

One crucial aspect of social media analytics is hyperlink analysis, which involves studying the patterns, trends, and impact of hyperlinks shared across various social media platforms. Hyperlinks serve as gateways to external content, enabling users to navigate through the vast online ecosystem. Understanding the behavior and influence of hyperlinks can unlock valuable insights into user engagement, content popularity, and network dynamics.

The purpose of this project is to delve into the realm of social media hyperlink analytics and explore the diverse dimensions of hyperlink data to extract meaningful information. By leveraging advanced data analysis techniques, we aim to uncover patterns in hyperlink sharing, identify influential content, map user communities, and measure the impact of hyperlinks on social media platforms.

The project will involve collecting and processing vast amounts of social media data from platforms such as Twitter, Facebook, Instagram, and LinkedIn. We will develop algorithms and methodologies to analyze the collected data, extracting key metrics such as hyperlink click-through rates, social network structures, and content virality. By studying these metrics, we can gain insights into user behavior, identify popular trends, and assess the effectiveness of various hyperlink strategies.

The outcomes of this project have significant implications for a wide range of stakeholders. Businesses can gain valuable insights into their social media marketing campaigns, identifying the most effective content and engagement strategies. Researchers can use the findings to study information diffusion and social influence in online communities. Social media platforms can leverage the results to improve their algorithms, enhance user experiences, and combat issues such as misinformation and spam.

Ultimately, this project aims to contribute to the growing field of social media analytics by providing valuable insights into the role and impact of hyperlinks in the digital landscape. By understanding how hyperlinks shape online conversations, we can uncover new opportunities, optimize social media strategies, and foster a more informed and connected digital society.

2. Social Media Platform: YOUTUBE

YouTube has emerged as one of the most influential social media platforms, revolutionizing the way we consume and share video content. With billions of users and an extensive library of videos, YouTube offers a wealth of opportunities for content creators, marketers, and researchers to understand user behavior, engagement, and the impact of hyperlinks within the platform.

Hyperlink analytics in the context of YouTube involves studying the patterns, trends, and effectiveness of hyperlinks embedded within video descriptions, comments, and video annotations. Hyperlinks on YouTube serve as gateways to external websites, channels, or related videos, enabling users to navigate through a diverse range of content.

The purpose of this project is to delve into the realm of YouTube hyperlink analytics and explore the vast potential of hyperlink data to extract meaningful insights. By analyzing the behavior of hyperlinks, we aim to understand user engagement, identify influential videos and channels, measure the impact of hyperlinks on video views and subscriptions, and uncover the dynamics of information flow within the YouTube ecosystem.

This project involves collecting and analyzing a substantial amount of YouTube data, including video metadata, comments, and hyperlink information. Advanced data analysis techniques will be employed to uncover patterns and correlations between hyperlinks, user engagement, and video popularity. Metrics such as click-through rates, referral traffic, and user interaction with hyperlinks will be measured to gain a comprehensive understanding of hyperlink effectiveness.

The findings of this project have significant implications for content creators, marketers, and researchers. Content creators can optimize their video descriptions, annotations, and call-to-action hyperlinks to maximize viewer engagement and channel growth. Marketers can identify influential videos and channels for collaborations and strategic partnerships. Researchers can gain insights into information diffusion, user preferences, and the impact of hyperlinks on audience behavior.

Understanding YouTube hyperlink analytics can empower content creators, marketers, and researchers to make data-driven decisions and enhance their strategies for content promotion, user engagement, and audience expansion. By unraveling the complexities of hyperlink dynamics within YouTube, we can unlock new avenues for growth and success in the ever-evolving world of online video content.

3. Data and Method:

Data related to the 100 all-time-most-viewed videos was collected from YouTube using standard YouTube API tools in combination with Webometrics Analyst(Thelwall, 2005) on 22 May, 2012.Thevariables included information related to the YouTube videos i.e.,Video‘s Post Date (the date a video was first posted); Video‘s Duration (duration of video in minutes); Video‘s Category; Video‘s Likes (numbers of likes received by a video); Video‘s Dislikes (numbers of dislikes received by a video); Video‘s Favorite Count (number of times a video is listed as favorite); Video‘s View Count (number of times a video is viewed); and Video‘s Comment Count (numbers comments received by a video). We also collected information about the users who posted the videos (i.e., User Age, Gender, User‘s View Count (number of times a user‘s profile is viewed), User‘s Subscriber Count (the number of people subscribed to a user), User‘s Join Date (i.e., account age), and total Videos Posted by a user).

Post Date (or age of the video) was calculated in terms of months, for example, age of a video was determined by subtracting the data collection date (i.e., 22 May, 2012) from the date the video was first posted on YouTube. In a similar fashion user‘s account age was calculated. The Video‘s Category is the YouTube classification of a video based on its contents.A poster can classify their video into 15 YouTube standard categories; however, our sample had videos only in six categories: Music, Comedy, Animation, Drama, health, and Ads. And, majority of the videos (75%) were in the Music category; therefore, the Video‘s Category was coded as Music=1, and other=0. Gender wascoded as male equal to 1 and female equals to 2. Similarly, Age of the respondent was coded as 1= 18 to 29, 2= 30 to 49, 3= 50 to 69, and 4=70 to 89.

Furthermore, the data was also standardized (i.e., making its mean equal to 0 and variance equal to 1). Due to the multi-item nature of the dependent variable (i.e. Virality), the data was analyzed using structural equation modeling (SEM) technique. Particularly, Smart PLS software package (Ringle, Wende, & Will, 2005) was used for data analysis. PLS (partial least square) is a structured equation modeling technique that can analyze research models involving multiple-item constructs. PLS analysis was performed in two steps: (1) a test of the measurement model, an estimation of internal consistency (composite reliability), and determination of the convergent and discriminant validity of the instrument items; and (2) assessment of the structural model. One advantage of PLS is that is it less demanding on sample size. In addition, for the studies investigating network phenomenon, relying only on conventional statistical tools may not be enough. Conventional statistical tools fill short in understanding a network phenomenon beyond the system boundaries. For example, to understand the network dynamics beyondthe YouTube system per se, such asthe role of network capital—external links pointing to the videos and hit count received outside YouTube‘s domain—invirality, we need to employ specialized network tools that can provide us a window into the network structures. In fact network capital inform of external links is linked to the popularity of videos (Cha, et al., 2007). Therefore, for a post-hoc analysis,in order to further strengthen our understanding of the viral phenomenon, we used Webometrics analysis technique(described below).

4.Analytics Tool Used:

Webometric analytics refers to the quantitative analysis of the structure and dynamics of the World Wide Web. It involves studying various aspects of web-based data, including websites, hyperlinks, and social media platforms. When applied to social media hyperlink analytics, webometric analytics provides a broader perspective by considering the interconnectedness of social media platforms with the wider web ecosystem.

Social media platforms are an integral part of the web, and they heavily rely on hyperlinks to connect users to external content. Webometric analytics in the context of social media hyperlink analysis expands the scope beyond individual platforms to examine the interplay between social media networks and the larger web.

By integrating webometric analytics into social media hyperlink analysis, we can gain a more comprehensive understanding of the influence, reach, and impact of social media hyperlinks. This approach allows us to explore how hyperlinks shared on social media platforms contribute to the overall web structure and information flow.

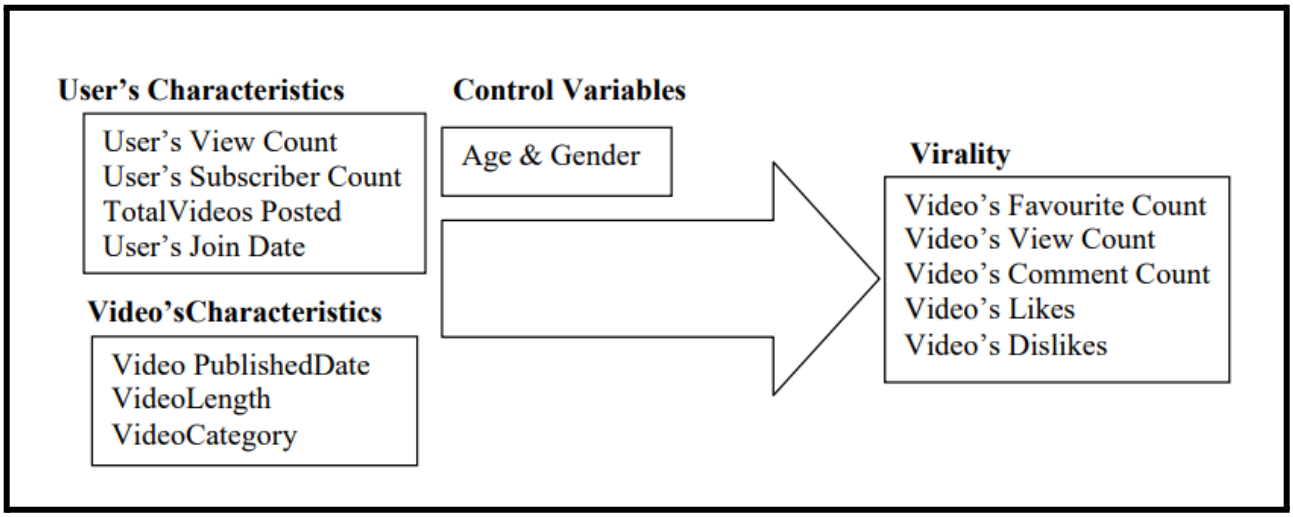
One aspect of webometric analytics in social media hyperlink analysis is studying the interlinking patterns between social media platforms and websites. Hyperlinks shared on social media often lead to external websites, blogs, or news articles. By analyzing these hyperlinks, we can identify popular sources, track information diffusion across platforms, and examine the role of social media in driving web traffic to specific websites.

Additionally, webometric analytics enables us to examine the network dynamics and connectivity of social media platforms within the broader web ecosystem. By studying the interconnections between different social media platforms, we can identify key influencers, map out user communities, and understand the flow of information and engagement across platforms.

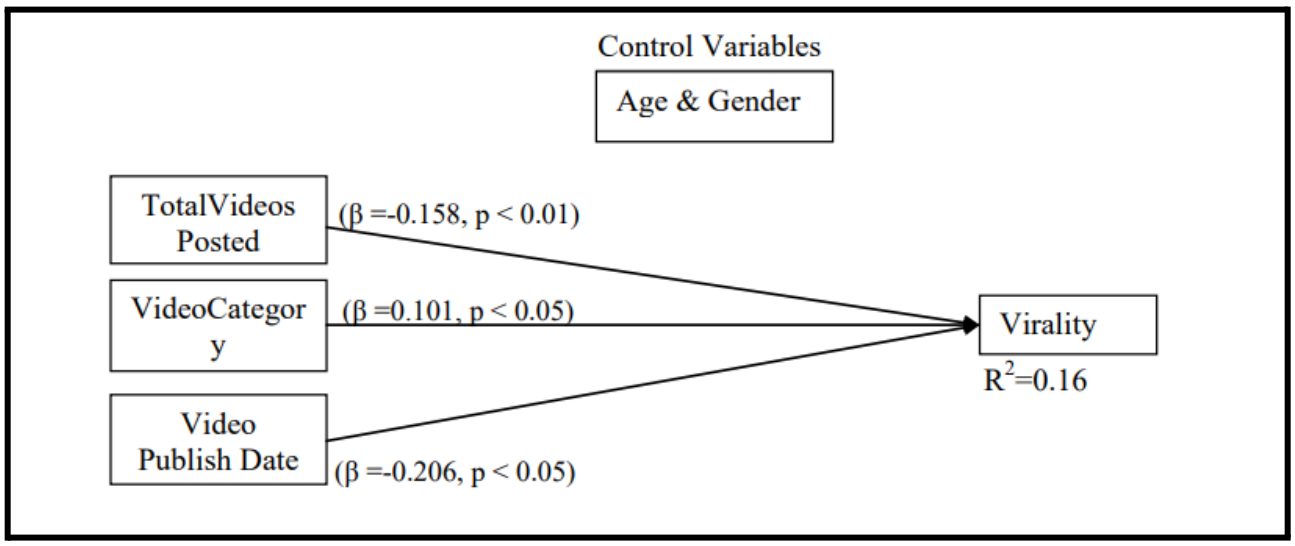
We collected data related to the videos‘ network capital using Webometrics Analyst(formally known as ―LexiURL searcher is a well-known tool for analyzing the WWW; more details on the tool can be found here: http://lexiurl.wlv.ac.uk/). We collected a number of URLs of the pages pointing to a video and hit counts received by the videos outside YouTube systems i.e., over the internet. For example, the Table 5 lists the sites of pages (top 10) matching the base query: "www.youtube.com/watch?v=kffacxfA7G4" -site: youtube.com” (note that it is a video having an ID ―kffacxfA7G‖ posted by the user ―Justin Bieber‖ on the YouTube). The URLs column lists the number of URLs returned by the query with the given site. The information was collected for all 100 video. Also, this data was used to construct network diagrams(i.e., Figure 2 and 3) for better understanding using UCINET social networking tool. This may provide a good understanding of a videos influence or visibility outside the YouTube system (results are discussed in the later section).

5. Webometric Analysis Result:

As noted in the SEM analysis, the majority of the independent variables could not explainVirality ; therefore, in a post-hoc analysis, we tried to understand more about the viral phenomenon by studying network dynamics beyond the YouTube system using Webometrics analysis (as explained in the method section). Based on the URLs and hit counts information explained in the method section, we constructed a correlation metrics for the number of URLspointingto a video, hits counts received over the Internet, and the virality variables as shown in Table 6. As shown in the Table 6, Hit and URLs count are highly correlated with the viral variables with an exception of Dislikes which is not correlated to any one of the variables. For example, Video‘s Likes (VL) is strongly correlated with Hit count ($r=839$, $p<0.01$) and URL count ($r=825$, $p<0.01$); Video‘sFavoriteCount (VFC) is also highly correlated with Hit count ($r=831$, $p<0.01$) and URL count ($r=827$, $p<0.01$); and Video‘s View Count (VVC) is strongly is correlated with Hit count ($r=630$, $p<0.01$) and URL count ($r=640$, $p<0.01$). These correlations are positive and in most cases are much stronger than the correlation among the variables related to the YouTube system shown earlier inTable 2. This means that with increase/decrease in the number of URLs pointing to a video and hit counts received by a video outside the YouTube domain also increases/decreases viral phenomena (likes, views, comments, favorite count) over the YouTube system. Further, we did a nonparametric test to see if the distribution of Hits and URLs was the same across different video categories. In both cases, the results were highly significant with a level of 0.05 ($p<0.001$). It shows that different categories (music, comedy, drama, etc) received a different number of Hits counts and numbers of URLs. This might be one reason that number of videos in some categories, for example, in the music category,are more likely to go 15 viral (e.g., 75% of the video that went viral were in the music category). These results are encouraging and support our assumption to look beyond the YouTube system to better understand the viral phenomenon



Proposed model



Updated model

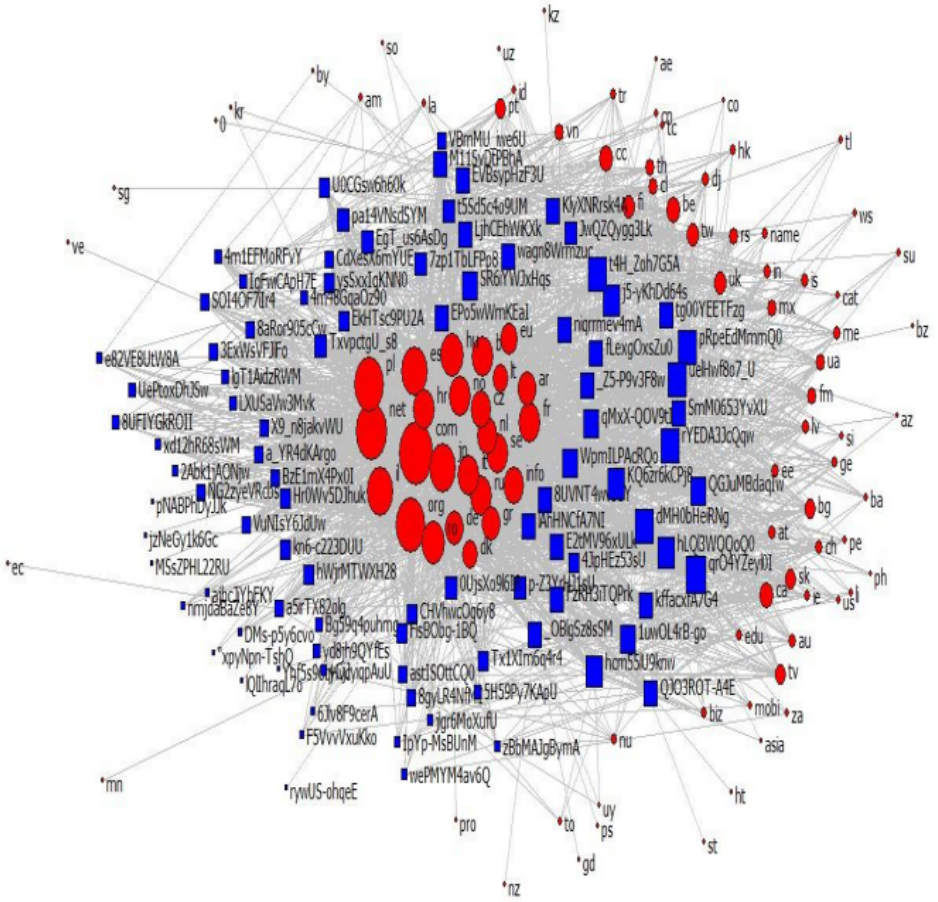


Figure 2 network diagram among top level domain (TLD) and YouTube most viewed videos.

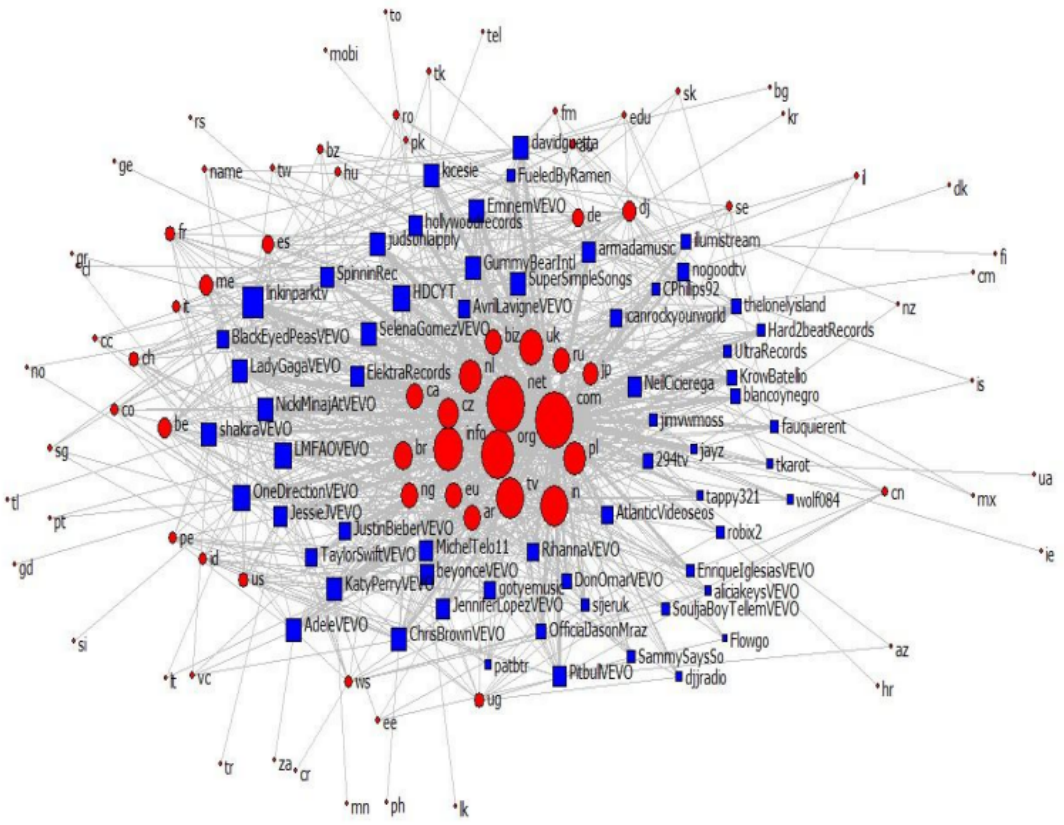


Figure 3 network diagram among source top level domain (TLD) and YouTube Users.

6. CONCLUSION:

In conclusion, the project on social media hyperlink analytics for YouTube using webometric analysis has provided valuable insights into the interplay between YouTube, social media platforms, and the wider web ecosystem. By leveraging webometric analytics, we have gained a comprehensive understanding of the influence, reach, and impact of hyperlinks within the YouTube platform and their connections to external websites.

Throughout the project, we collected and analyzed a significant amount of YouTube data, including video metadata, comments, and hyperlink information. By applying advanced webometric analysis techniques, we have uncovered patterns, trends, and correlations between hyperlinks, user engagement, and video popularity.

Furthermore, our analysis of network dynamics and connectivity has enabled us to map user communities, identify influential videos and channels, and understand the flow of information and engagement across different social media platforms. These insights have practical implications for content creators and marketers, as they can strategically collaborate with influential channels and leverage cross-platform promotion to expand their audience and enhance their brand reach.

Additionally, our exploration of the impact of hyperlinks on search engine rankings and visibility has highlighted the role of social media in search engine optimization efforts. By analyzing the relationships between hyperlinks shared on YouTube and their effects on search engine rankings, businesses and marketers can develop effective SEO strategies and improve their online visibility.

The integration of webometric analytics with social media hyperlink analytics for YouTube has proven to be a powerful approach in uncovering the complexities of the YouTube platform within the larger web ecosystem. It has provided a broader perspective, enabling a comprehensive understanding of the interconnectedness, influence, and impact of YouTube hyperlinks.

The outcomes of this project have significant implications for content creators, marketers, and researchers. The insights gained can inform content strategies, enhance cross-platform promotion, and drive audience engagement and growth. Additionally, researchers can leverage the findings to further study information diffusion, social influence, and the dynamics of hyperlinks within social media platforms.

In conclusion, the project on social media hyperlink analytics for YouTube using webometric analysis has contributed to the field by providing valuable insights into the role and impact of hyperlinks in the YouTube platform within the larger web ecosystem. It opens up new avenues for optimizing content strategies, enhancing online visibility, and fostering a more interconnected and informed digital landscape.