Early Viewers or Followers: A Mathematical Model for YouTube Viewers Categorization

Niyati Aggrawal and Anuja Arora
Department of Computer Science Engineering, Jaypee Institute of Information Technology, Noida, India

Adarsh Anand
Department of Operational Research, University of Delhi, New Delhi, India

Yogesh K. Dwivedi
School of Management, Swansea University, Swansea, UK

***Abstract*—**

Purpose: Since the emergence of video sharing sites from early 2005, YouTube has been pioneering in its performance and holds the largest share of Internet traffic. YouTube plays a significant role in popularizing information on social network. For all social media sites, viewership is an important and vital component to measure diffusion on a video sharing site, which is defined in terms of the number of view counts. In the era of social media marketing, companies demand an efficient system that can predict the popularity of video in advance. Diffusion prediction of video can help marketing firms and brand companies to inflate traffic and helps the firm in generating revenue.

Methodology: In the present work, viewership is studied as an important diffusion affecting parameter pertaining to YouTube videos. Primarily, a mathematical diffusion models proposed to predict YouTube video diffusion based on the varying situations of viewership. The proposal segregates the total number of viewers into two classes - Neoterics Viewers i.e. Viewers those viewing a video on a direct basis and Followers: Viewers those watching under the influence.

Findings: The approach is supplemented with numerical illustration done on the real YouTube data set. Results prove that the proposed approach contributes significantly to predict viewership of video. The proposed Model brings predicted viewership and its classification highly close to the true value.

Originality/value: Thereby, a behavioral rationale for the modeling and quantification is offered in terms of the two varied and yet connected classes of viewers- “Neoterics” and “Followers”.

Keywords: Virality,Information Diffusion, Popularity Dynamics, Viewership, YouTube, Prediction

1. INTRODUCTION

Online social network is now everywhere. It has brought people with the same interest under one canopy irrespective of their demography i.e. people of different age groups or different origins having a similar interest towards social media posting content are now connected. The process has made information sharing a vital part of human life. Nowadays, people react in the form of share, comment, like on the posted contenton social sites. Indeed, the identification of videos that will be liked and shared the most has become a crucial research task. Basically, two fundamental methods have been investigated by researchers to improve the content reach among the mass and maximize the effect. First is the emergence of content on social networking sites and second is directed through any outside source such as word-of-mouth or survey. In a recent study, 64% of marketing executives indicated that they believe word of mouth is the most effective form of marketing [53].Researchers who are working to predict the popularity of social networking sites have used the statistical data of users’ actions on posted content. The notion of predicting video popularity is to model the behavior of viewers towards a video.

Predicting and understanding the popularity is useful from two-fold perspectives- First, it leads to the generation of more internet traffic and second it has adirect economic impact. Researchers have introduced various methods for popularity prediction[49][17][23][55].The first work towards understanding the popularity of YouTube videos is done by Chatzopoulou et. al. in 2010 [17]and in their work they have found that four metrics- view count, comments, rating, and favorites are having a high correlation with video popularity. FurtherGursun et. al.[23]have worked on change classification patterns of YouTube videos and classified videos on two categories- Frequently access videos and rarely access videos. They have applied SVM, Autoregressive Moving Average, and hierarchical Clustering to achieve the outcome [23]. User contextual information has also been used for popularity forecasting purposes [55][51].

YouTube is considered as one of the most popular social media to work on these video diffusion issues by researcher community[1][2][29] as it drives a significant amount of web traffic and can be used by a business community to advertise their products in the form of slides, animated videos or in some other pictorial form. Circulating these advertisements on YouTube provides additional social signals for search and each video page can be optimized to enhance the ability for driving web traffic back to its site. Although, largest online social networks in China [55], Facebook [7], Amazon [7], Instagram [6], and Twitter [6]are also used by internet researchers for popularity forecasting.

YouTube is considered as a well-known name in the field of video sharing sites [57]. According to Unmetric[1] analytic report approximately every single minute three hundred videos are uploaded to YouTube which is available to more than 1 billion YouTube users in 75 countries and in 61 languages. Since its arrival, it has become revolutionary in terms of the videos submitted to it as well in terms of the rate of subscription. Toboola[2] is a well-known promotional site that optimizes conversions on videos, drives growth, promotes online content and increases traffic. Taboola claims that by 2020, YouTube videos have accounted for 69% of all consumer Internet traffic. None of its rivals (Dailymotion.com[3], Vimeo.com[4] or flickr.com[5], etc.) have reached nearby YouTube and its popularity has been growing steadily.

On YouTube, various kinds of videos originate every day which can either go viral [58] and can inspire heated discussion or die out immediately which has diverted our attention towards the investigation of diffusion dynamics. From a video uploader's point of view, it is important to identify the number of viewers seeing the video through external influence and count of users viewing the video under somebody else's influence. On one hand, this kind of study can help the organizations to influence the target audience which can have a direct impact on the overall revenue generated by the company. Also, it may help in planning to discard inimical information early such as rumors and inauthentic news which may spread unnecessary annoyance.

YouTube video preference primarily depends on two factors: 1) Video characteristics i.e. Video Title, Video Length, Video Age, and Video uploader; 2) User Generated Content (UGC) i.e. view count, like count, comment count, subscribers count or dislike count [36]. In 2019, Aggarwal et. al. tried to correlateVideo characteristics and UGC[8]. In their work, they came up with some important correlation insights of video such as all view metrics are directly proportionate to engagement metrics, video virality is inversely proportionate to video age [20], etc. They have concluded the YouTube viewer takes the decisions to prefer the particular video on the basis of characteristics of video (i.e. video length, video age, video up-loader or video title) or on the basis of data/ statistics generated by the other users (i.e. video view count, like count, comment count, subscribers count or dislike count).Video length and title are important factors that attract viewers to prefer a particular video. Videos are searched by its title and users prefer the most matched video. Another attribute of likeliness is the shorter length of a video [8].

In this paper, a mathematical View-Count Model (VCM) is conceptualized that utilizes daily view count dynamics (DVCD) of YouTube videos for viewers' categorization. Studied literature showcased that popularity dynamics of YouTube videos adapt lognormal distribution. Borghol et. al. in 2011[12]and Kamiyama et. al.in 2019[29] have found similar findings in their work. The lognormal distribution is applicable on YouTube videos whenever a user views the video because in lognormal distribution log(x) exists when x is positive. In the case of YouTube, it will never be negative. Few models are designed by researchers for popularity distribution but categorization of viewers is not attempted in existing researchers. In popularity distribution work, a Simple Time-series model i.e. Simple Multiplicative process (MPP) has been used [38]. In most recent work, Kamiyama et. al.[29] used MPP to model the view dynamics of YouTube video for reproducing popularity distribution. Still, the research gap exists for video popularity prediction and viewers categorization. With this aim, a mathematical modeling framework is proposed to categorize viewers (early viewers and followers) and to get exact view count for each categorization for a specific video.

[1] https://unmetric.com/youtube-analytics
[2] http://go.taboola.com/promote_videos_in/
[3] http://www.dailymotion.com/in
[4] http://www.vimeo.com
[5] https://www.flickr.com/

In this paper, keeping in mind the ideologies reported above, an initial assumption has been made to observe the characteristics and distinction of the viewers that might be viewing a specific video. VCM model computes the precise count of viewers in two categories and these categories are referred to Neoterics and Followers. Those who watch a video on a direct basis (through external influence) are 'Neoterics' users and those individuals who watch a video under someone else's influence; are termed as 'Followers'. Various performance measures [4]5][24]such as Sum Squared Error (SSE), Mean Square Error (MSE), Root Mean Square Deviation(RMSE)and Co-efficient of determination(R-Square) are used to validate the outcome of the proposed VCM model. MAPE is used to evaluate the level of forecasting.
The contribution of the paper is summarized as the following two points:

- A mathematical modeling framework that utilizes dynamic daily view count (DDVC) of various YouTube videos to predict the video view count has been proposed.
- The proposed model categorizes the Video View Count in two categories: Neoterics and Followers which gives accurate information of mode of video diffusion. Even this classification can be used to get insights about organic and paid viral videos.

The article is chronically arranged as follows: Section 2comprises of the related literature and background related to research done in the field of information diffusion. Section 3 discusses the preliminary of the proposed video diffusion model, notations and formulation of the proposed modeling framework which is followed by detailed descriptions about datasets in section 4. Data analysis and results are depicted in section 5.Experimental interpretations of the proposed modeling framework have been highlighted in section 6. Finally, concluding remarks and future scope are discussed in sections 7.

2. LITERATURE REVIEW

It is a well-known fact that diffusion is a social process and the social strength of peers led to adoption by many potential adopters. Predictive models aim to formulate models to understand the diffusion and adoption process which depends on social strength. The work presented in this paper has its roots arising from theory of adoption and diffusion of new ideas or new products in a social system that has been discussed at length by various researchers- Bass [10], Mahajan [35], Anand et al. [5], Dwivedi et al. [19], Asch [3], William et. al.[54] and Rogers [43]. There are various fields such as Agriculture, Medical, health [50], Marketing [31],Virus propagation, Sociology, Communication Technology, etc. on which these models had been applied by various researchers to examine the diffusion of information over large networks. Some existing predominating models based on diffusion process are as follows:

- **Threshold Model [21]:** The base concept of the Threshold models to choose a value which is a proportion of other peoples who agreed on one choice (threshold value) before a given actor agrees on the same decision.
- **Independent cascade Model [30]:** This model examines the social influence or how the behavior of others affects the overall rating of his specific content. On social platforms, it is very common to reuse the content of others (friends of mutual friends). People decide to take on the content/behaviors of their friends based on some weightage / closeness that they give to their relation.
- **Epidemic Model [28]:** Epidemic models are used to study virus propagation in any individual including human-being, animals, computers or plants. Under these models, three kinds of individuals categories can be examined such as suspected with viruses, infected with a virus or recovered from the virus.
- **Critical Mass Model [37]**: This model is popular to study the mass/crowd behavior and the collective deed/action on social platforms. This model can be used to study the various real-time applications such as social platforms (in understanding the diffusion of ideas and innovations), political sciences (diffusion of market-oriented strategies), crime sectors (useful in explaining the affinity for crime rates to knowledge "explosions "and "arrests ").

Out of all above-mentioned models our current proposal lies in the epidemic model category. In recent years, content virality has gained huge attention due to the increasing amount of users' involvement in the social web, which has overwhelmed the users and users are not able to interpret the most relevant and desired option on social media. In social media literature, limited research work is available in the context of analyzing content virality. This section summarizes the works that are most representative and relevant to the study. The focus is primarily on research work which is going on all around information virality/information diffusion. Most of the researchers have used Facebook and Twitter data to visualize the behavior of content popularity in terms of its viewership and the only handful of research is being done on YouTube dataset to predict the behavior of video popularity based on viewership. Existing work is focused on the distribution of YouTube videos statistics,

YouTube video popularity prediction, and virality over this video sharing site. In the field of marketing, the diffusion process narrates the manner in which the new product penetrates the market. Moving with this definition of diffusion, our inclination is to understand the diffusion process of a video being posted on YouTube.

The timeline view of research work performed in this direction is shown in Figure 1. In 2007, Cha et al. [14] took user-generated content site (UGC) YouTube and non-user generated content site Daum (Korea) into consideration for the trace-driven analysis of UGC and non-UGC video. Paolillo [39] scrutinized the social structure of YouTube based on friendly relations with their corresponding tags applied while video upload. The experiments are performed on users profile dataset which contains the profile of users, their friends' comments on the video, video details (in video references) and Video author[39].
Rotman and Preece [42] examine the growth of the YouTube online community through the eyes of YouTube users like bloggers. They selected the specific subgroup of video loggers (users who are chatty, energetic, opinionated i.e. involved in the idea of YouTube community) [42]. Another study by Davidson et al. [18] has discussedthe recommendation system for YouTube videos. In 2014, Khan and Vong [20] assembled and verified an empirical model to understand the relationship among users' social and non-social capital, video characteristics, external network capital (in-links and hit counts), and virality of YouTube videos. In this work, they explored the different categories of videos and not worked on the contents of the videos intrinsically whereas video content plays an important role in the viral phenomenon.

Vaish et al. [52] grasped the conclusion that virality grows and falls exponentially and virality follows similar patterns with time. They used the popularity variable YouTube dataset, simple pathogenic epidemic model, conventional quantitative asset valuation method, hybrid valuation method, to quantify virality (viral index formula) of videos. Topical research by Cheng et al [16] presents a systematic and in-depth measurement study on the statistics of YouTube videos. They found that YouTube videos have noticeably different statistics compared to traditional streaming videos, ranging from the length and access pattern to their growth trend and active life span. Popularity has one more dimension; it has been defined in terms of the view-counts that any video generates. The more the view-count implies more popular the video is. Moreover, a study by Lange et al [32] shows how circulating and sharing videos reflect different social relationships amongst participants.

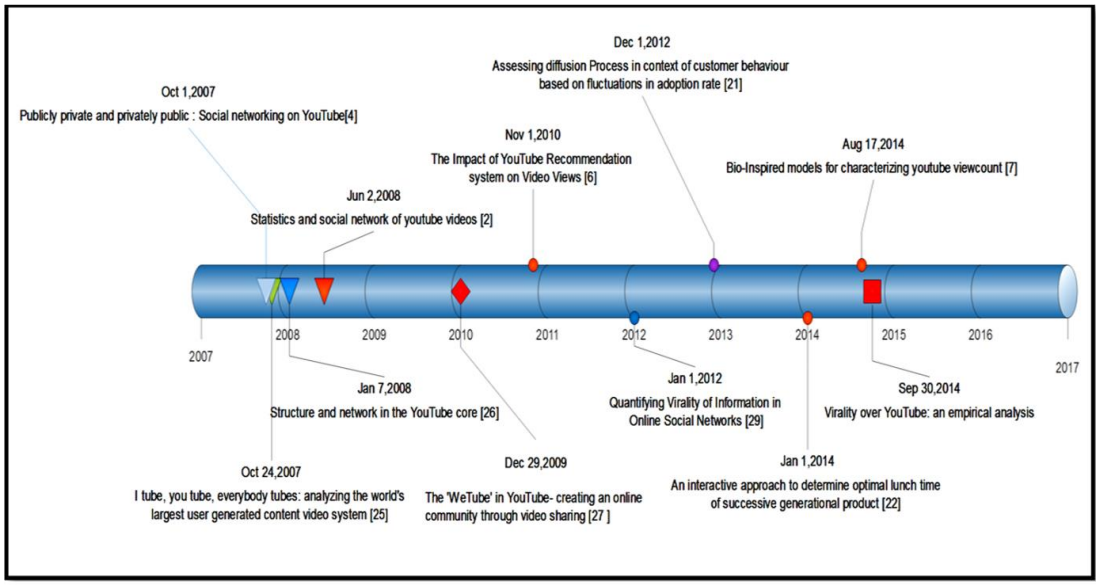


Figure 1: Timeline review of research based on videos posted on YouTube

Of late, the research in this theme of video sharing has been centered on the models proposing early-stage predictions for popularity in the future [40]. Furthermore, as discussed by Zhou et al [56]; the understanding of how certain features in a video drive the viewers is useful for creating a strategy to drive any video's popularity. They have proposed an approach to measure the data sets crawled from YouTube and its related recommendations, wherein they claim that despite the fact that the YouTube video search is the number 1 source of views in aggregation, the related video recommendation is the main source of views and hence the major reason of view counts. But these studies have not talked anything about popularity dynamics using attributes like view-counts and its related modeling.

Research in the area of view count characterization is not new butthe only handful of studies exists that have talked about the mathematical modeling concept for viewer categorization and viewership prediction. Like a recent work done by Richier et al.[41] describes some of the most typical behavior of the view-count of videos on YouTube. They provided in-depth analysis and used models that capture the key properties of the observed popularity dynamics. Further, they have claimed to match the observed videos view counts with one of several dynamic models. To select candidates for these models, they have made use of bio-inspired dynamics. For doing so, they have claimed that the propagation of content on YouTube has a strong similarity with the temporal behavior of an infectious disease, which is a classical topic in mathematical biology [9, 34, 40].Such models of disease spread have already been used in order to model the spread of viruses in computer networks [15, 22]. They have been also used in marketing for capturing the life cycle dynamics of a new product [4] [11] [24] [48].Study by Aggarwal et al. [2] describes an important view-count based modeling framework in the context of a dynamic internet market. The authors proposed a ranking methodology to find out the best model out of the three dynamic internet marketing models. Yet another work by Bisht et.al [11], describes about utility of Interpretive Structural Modeling approach for understanding popularity dynamics for YouTube videos.

Recent advancement in the information diffusion modeling has been given by Irshad et al [24]; wherein; the authors modeled the active life span of You Tube videos depending upon the change in viewership rate. In another work by Irshad et al [25] the authors described a modeling framework to study the popularity dynamics based on YouTube viewers and Subscribers. But the current work has its roots embedded with the work of Aggrawal et al [1], wherein an approach to differentiate the viewers based on the time frame they take to watch a video has been presented.For a video, the characterization of viewers/participants is very important [24]. This is so because a system recommends personalized sets of videos to users based on their activity on a particular platform. Thereby making it important to distinguish the users who are directly being influenced by external sources and engage themselves in watching a particular video to users whose behavior depends upon the existing viewer's recommendation. Therefore, a video site must know which and what type of participants are having a look at their video [13, 49]. None of the above described models defined a predictive analysis model for video virality/ influence propagation nor studied the video influencing process which led to the formulation of this study. With the similitude that can be established between the innovation diffusion model as given by Bass [10] and virality prediction model is given by Aggrawal et al. [1]; an approach to quantify the count of Neoterics and Followers has been formulated.

3. VCM - A MATHEMATICAL VIDEO VIEWERSHIP PREDICTIVE ANALYSIS MODEL

User's decision to adopt a new product or watch a particular video clip often depends on the distribution of similar choices the individual observes among its peers (be their friends, colleagues, or acquaintances)[47]. It may also be an artifact of simple learning processes, where the chance that an individual learns about a new offering or its benefit from their peers is increasing. For instance, decisions regarding whether to go to a particular movie or restaurant, or whether to watch a particular video, provide examples of situations in which information learned through friends and their behavior are important. Of course, there are many other potential channels by which peer decisions may have a significant impact on an individual's behavior. The theory of timing of the initial purchase of a new consumer product has been a very popular and core area of research in Innovation diffusion modeling in marketing [24, 4, 5, 31, 19, 35, 43, 44, 45, 48].The VCM model blends marketing analogy and the mathematical model proposed by Aggrawal et al [1]. VCM framework characterizes the viewership process in terms of the video view count. Basically, this model is established to derive a relationship connecting the innovation diffusion process with the viewership process. In marketing science, viewers are basically categorized into two differing groups- Early Viewers and Followers. This work details an impressive approach which effectively and accurately categorizes viewers in thefollowing two categories:

Neoterics: Some individuals might watch a video independently of the decisions of other individuals in a social system and rather a handful of them also have the influencing power to influence others to have a look at what they are liking. We shall refer to these individuals as *Neoterics*.

DEFINITION1: (Neoterics) A group of the user who is first to view the video and then trigger its diffusion process.

Followers: Apart from Neoterics, viewers are influenced by the pressure of the social system and differ in timing of viewership; we can call such viewers as *followers.*

DEFINITION2: (Followers) Pool of users who view the video through WOM (word of mouth) communication received from others and then triggers its information diffusion process.

Definitions of symbols are summarized in Table 1 which has been used to construct the VCM Model.

Table 1: Definition of Symbols

Symbol	Description
v_1	rate of viewership forNeoterics
v_2	rate of viewership for Followers
N	Internet market size.
T	Time factor
V(t)	Cumulative number of viewers till time‘t’.
f(t)	Likelihood of viewership at time‘t’.
F(t)	Cumulative likelihood of viewership by time‘t’.

The VCM model is designed by considering a few pre-assumption prior to applying. These pre-assumptions are basically some present model constraints which should be catered by researchers in the near future [1]:

- Potential viewers watch a specific video precisely in two conditions, Influenced by external sources or internal sources.
- Potential viewers can watch a particular video only once. Repeat viewers are not under consideration while forming of VCM model
- The diffusion process for a video is a binary process.
- The characteristics of a video that is under study and its perception do not change.

Aggrawal et. al. revealed that analogous modeling framework given by Bass [10] can be used for early viewers and followers classification [1]. Moreover, this work classification has been done based on time i.e. viewers are classified on the basis of time to watch a video. Our model is basically an extension of the model proposed by Aggrawal et. al.[1]. The base concept of theBass model is to investigate time factor and categorize users in innovators and imitators [10] based on their content access time. In our work, we can also categorize users on the tendencies of the Bass model

To estimate the viewership of YouTube videos in a generic manner, it is desirable to find out the association of two types of viewers that are contributing to the overall view count. It is assumed that overall viewership is initiated by a certain number of viewers at the time of model initialization after the launch of video.Further, the rate of viewing any video at a given time comprises of two components that administrate the viewing process; the first factor constitute the videos watched through external influence with an impact rate v_1 and the second factor represents the additional number of viewers who watch a video under the influence of peers with influence rate v_2.

Let N be the pre-determined pool of viewers / Market size those will watch the video. We aim to make the viewership prediction i.e. number of new viewers v_t at a particular time$t$. v_1 and v_2 are viewers those will watch the video as Neoterics and Followers respectively. Let $V(t)$ be the cumulative viewers. Equation 1 is a differential equation that describes the viewership prediction model. In the model, followers v_2 are multiplied by people who already have seen the video i.e. $V(t)/N$.

$$v_t = \frac{dV(t)}{dt} = (v_1 + v_2 \frac{V(t)}{N})(N - V(t)) \quad (1)$$

Where, v_1 represents the fraction of all viewers those watch video on their own i.e. Neoterics. Neoteric users’ decision to watch a particular video is either through self influence or direct advertisement policies. The product v_2/N times $V(T)$ reflect the pressure operating on Followers as the number of Neoterics increases in the system. After solving the equation (1) the closed-form solution is obtained as

$$V(t) = N\left(\frac{1-e^{-(v_1+v_2)t}}{1+\frac{v_2}{v_1}e^{-(v_1+v_2)t}}\right) \quad (2)$$

Therefore, if N is the total number of views of a video, then the cumulative number of viewers who have watched by time t i.e. $V(t)$ can be rewritten as given in equation (2)

$$V(t) = N \cdot F(t) \quad (3)$$

The behavior of potential viewers as described by above presented model will surely carry researchers' attention as its results are uniformly exquisite same as Marketing Management Model proposed by Bass [10]. Even, our proposed information diffusion study is identical to the Software Reliability Growth Model given by Kapur and Garg [26] in the field of software engineering.Moreover, if in equation (2) we substitute $b = v_1 + v_2$ and $\beta = (v_2 / v_1)$ then theabove described model reduces to the prototype as given by Kapur et al. [27].It is very interesting to note the behavioral rationale for the aforesaid number of viewers calculated based on view count. View-count comprises both Neoterics and followers and the important distinction between these two user categories is their video watching influence based on timings.
With all points leading to effective marketing, brands use strategies for a particular period to create some great promotional material [50]. Therefore, a video is generally created in such a manner that it could attract a large number of viewers. The group of Neoterics that we want to discuss here caters to people of this category. On the contrary part many times despite all the effort, video attracts few viewers i.e. less engagement.

Thereby, we define, $f(t)$ as the likelihood of viewers at time $t$and N to be the total number of viewers during the period for which density function was constructed. Also, we let $F(t)$ to be the fraction of Netizens covered by time$t$ or cumulative likelihood of viewership by time t. The likelihood of viewership at a given time$t$using equation (1) and the result obtained through equation (3) can be created as follows:
From equation (3), $F(t)$ can be written as$F(t) = {}^{V(t)}/_{N}$. Therefore, equation (1) can be written in terms of $F(t)$ as follows.

$$\frac{dF(t)}{dt} = v_1\left[1-F(t)\right] + v_2F(t)\left[1-F(t)\right] \quad (4)$$

Equation 4 classifies viewership in two fractions a and b. where,$a = v_1[1 - F(t)]$, Fraction of Neoterics and $b = v_2F(t)[1 - F(t)]$, Fraction of Followers in complete viewership N. Equation (4) is transformed in terms of $f(t)$:

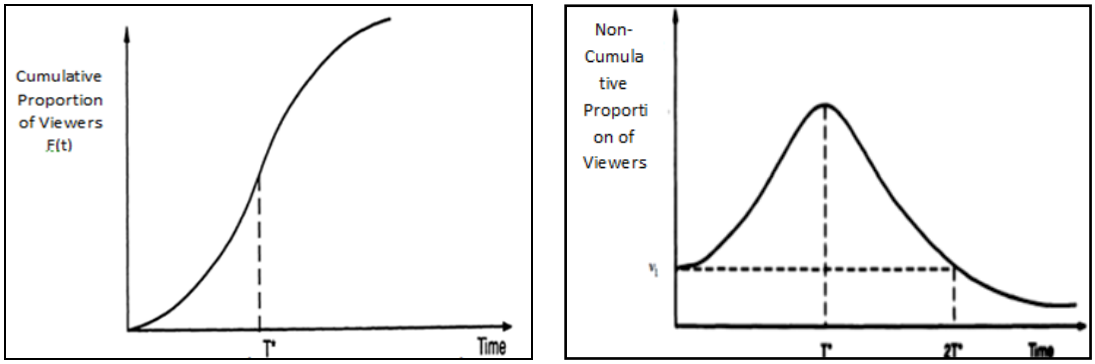
$$f(t) = \frac{dF(t)}{dt} = \left[v_1 + v_2F(t)\right]\left[1-F(t)\right] \quad (5)$$

The solution to equation (5) yields s-shaped cumulative viewership distribution shown in Figure 2(a) and can be given as

$$F(t) = \left(\frac{1-e^{-(v_1+v_2)t}}{1+\frac{v_2}{v_1}e^{-(v_1+v_2)t}}\right) \quad (6)$$

Further, the differentiation of $F(t)$ gives the non-cumulative viewers' distribution as shown in Figure 2(b) which represents the stated view process as:

$$f(t) = \frac{v_1(v_1+v_2)^2 e^{-(v_1+v_2)t}}{(v_1+v_2 e^{-(v_1+v_2)t})^2} \quad (7)$$



(a) Cumulative Proportion of Viewers F(t) (b) Non-Cumulative Proportion of Viewers f(t)

Figure 2: Viewership Distribution

It is clear in Figure 2(b) that the curve achieves its peak at $f(T^*)$ or $F(T^*)$ at time T^* where:

$$T^* = -\frac{1}{(v_1+v_2)} \ln\left(\frac{v_1}{v_2}\right), \quad (8)$$

$$F(T^*) = \frac{1}{2} - \frac{v_1}{2v_2} \quad (9)$$

$$f(T^*) = \frac{1}{4v_2}(v_1+v_2)^2 \quad (10)$$

It can be observed in Figure 2 (b) that the noncumulative viewership distribution is symmetric concerning time. It can be further shown that $f(t=0) = f(t=2T^*) = v_1$, that is, the proportion of non-cumulative viewer's distribution is symmetric with respect to time around the peak time $T*$ up to $2T*$.

Further, Figure 3 represents the non-cumulative view count curve for both categories of viewers. As stated earlier, a or $v_1\left[1-F\left(t\right)\right]$ in equation (4) represents the Neoterics i.e. viewers watch the video through external influence. On the other hand, term b or $v_2F\left(t\right)\left[1-F\left(t\right)\right]$ in equation (4) represents the followers i.e. those viewers watch the video under someone's influence and thereby their name.

Indeed, Figure 3 depicts the varying behaviour of self-motivated to watch and followers viewers. Even, the interpretation of Figure 3 is very interesting to note. It is observable in the figure that in the beginning phase of the video diffusion process that there are some viewers present on the initial basis those watching video through external influence and in the later stage of video life-cycle; followers come into action and take charge of popularising the video.

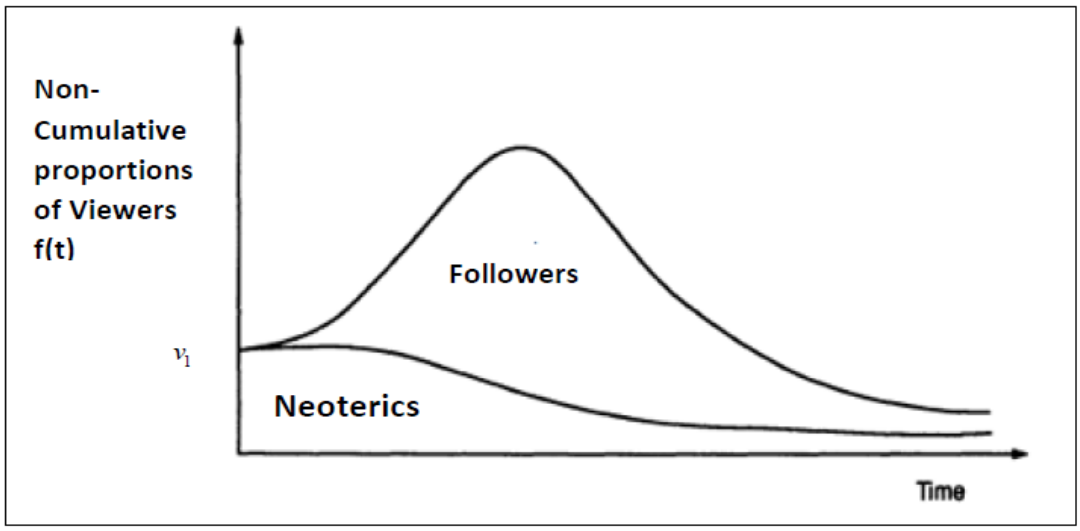


Figure 3: View Count Comparison of Neoterics and Followers

Since we want to distinguish various types of viewers amongst the Netizens, we propose a method to quantify the number of different viewers present in the social system. As assumed, $F(t)$ is the cumulative number of viewers by time 't' . According to the proposed model, there are two categories of viewers, $F(t)$ can be assumed to be comprised of two components; $F_1(t)$ corresponding toNeoterics and $F_2(t)$ corresponding to followers where, the total number of front runners, i.e. Neoterics $F_1(t)$ between any two time periods, say t_0 and t_F $(t_F > t_0)$ is given by

$$F_1(t) = v_1 \int_{t_0}^{t_F} [1 - F(t)]dt \qquad (11)$$

Substituting the mathematical form $F(t)$ is given by equation (6), hence $F_1(t)$ can be inferred as

$$F_1(t) = v_1 \int_{t_0}^{t_F} \left[1 - \frac{1 - e^{-(v_1+v_2)t}}{1 + \frac{v_2}{v_1} e^{-(v_1+v_2)t}} \right] dt$$

This after integrating gives,

$$F_1(t) = \frac{v_1}{v_2} \ln \left[\frac{v_1 + v_2 e^{-(v_1+v_2)t_0}}{v_1 + v_2 e^{-(v_1+v_2)t_F}} \right]$$

Substitution of $t_0 = 0$, $t_F = t$ in the above equation yields:

$$F_1(t) = \frac{v_1}{v_2} \ln \left[\frac{1 + \frac{v_2}{v_1}}{1 + \frac{v_2}{v_1} e^{-(v_1+v_2)t}} \right] \qquad (12)$$

The proportion of views by followers can be given by making use of equation (12) and subtracting it from 1, i.e. $F_2(t) = 1 - F_1(t)$, therefore;

$$F_2(t) = \left[1 - \frac{v_1}{v_2} \ln \left[\frac{1 + \frac{v_2}{v_1}}{1 + \frac{v_2}{v_1} e^{-(v_1+v_2)t}} \right] \right] \qquad (13)$$

The expressions obtained in equation (12) and (13) are very important to interpret. After multiplication with *N*, they give the required and exact count of the number of viewers falling under each category. For any video sharing site, this can help to know how much effort one is putting up and how much more shall be required to diffuse their positioning in the internet market.

4. DESCRIPTION OF DATA SET

YouTube is one of the biggest brand ambassadors for video sharing sites. It contains almost every type of online video content including various categories - Music, Sports, Gaming, Films, TV Shows, News, Live, Spotlight, 360° Videos, etc. [14]. All the videos have their own metrics based on user engagements on YouTube. YouTube metrics are basically a statistical measure and these metrics regularly update whenever gets user's attention. Few well-known metrics include View Count, Like Count, Dislike Count, Share Count, Comment Count, etc.
In order to understand the data and its time changing nature, we have extracted view count for 30 days period of thevarying category of YouTube videos. By taking the varying category of videos, we can uniquely determine the behaviour of the proposed prediction model on all categories. Even, we have a targeted number of videos ranges from 37 to 120 for different categories. Video streaming data of view count of all category videos are extracted using YouTube Data API v3.
VCM Model is designed based on the view count metric. So, in our work exclusively view count has been extracted using API whereas this data API can be used to extract numerous YouTube Video features. Moreover, varying features such as length, age of video are taken into consideration while data capturing processes to validate the compatibility and effectiveness of the proposed VCM Model. The data descriptive details are shown in Table 2. It is observable in the table that video length varies from 3-10 minutes, average video age ranges from 15- 75 days, and view count varies from 5000-12000.
We regard the readers and make them comfortable with the VCM learning process. Therefore, the detailed experimental evaluation for 6 varying categories of videos is presented in detail in further sections. The period for which view count data is collected is shown in Table 3.

Table 2: Category and characterization of You Tube Videos

Category	#videos	Video Length (in seconds)			Mean Video Age (in days)	Video View Count		
		Min	Max	Average		Min	Max	Average
Music	~120	5	3307	219	612	1	5527023	8305
Entertainment	~70	2	1732	245	609	0	2046258	11545
Comedy	~50	5	1442	195	605	2	1041238	6243
Sports	~45	3	851	166	618	2	1862136	5412
Movies	~62	2	2687	357	654	3	432745	7154
News & Politics	~37	258	754	604	628	2	625753	12352

Table 3 is a detailed description of datasets (DS) on which model performance outcome has been evaluated. In the taken dataset, DS-I is the temporal dynamics of view counts of music video for a period of 27 days, DS-II is for Song category for 26 days, DS-III is drama category video for 20 Days, DS-IV is again Drama category video and data collected for 15 days, DS-V is Song category video for 37 days and last DS-VI isMovie Trailer comprises of cumulative view counts for a period of 35 days. The time period of starting of collection of view counts is described in Table 3, which is as follows:

Table 3: Descriptive Details of Demonstrated Dataset

Data Set	Video Category	Period for which view count is collected
DS-I	Musical	16th September 2018 to 30th September 2018
DS-II	Song	4th September 2018 to 30th September 2018
DS-III	Drama	10st September 2018 to 30th September 2018
DS-IV	Drama	15th September 2018 to 30th September 2018
DS-V	Song	23th August 2018 to 30th September 2018
DS-VI	Movie Trailer	25th August 2018 to 30th September 2018

5. NUMERICAL RESULTS

To evaluate model performance and effectiveness, we carried out experiments on real datasets. The experiments were designed to answer the following research questions (RQs):

- RQ1: Is it possible to predict the exact count for Neoterics and Followers and thereby visualize the video diffusion pattern?
- RQ2: How well the model can predict the video diffusion pattern?
- RQ3: Can something be highlighted about the virality of a specific video?

Research question 1 is answered through section 5.1 which shows experiments performed on some sample videos for result demonstration. Video diffusion pattern has been measured through the proposed mathematical model and the predicted values are listed in Table 4(a) -4(f) and further visualization of the diffusion pattern is depicted in Figure 4 (a-f). Research question 2 is answered in section 5.2 which talks about various performance measure metrics. Research question 3 is answered in the interpretation section, i.e. in section 6.

5.1. EXPERIMENTAL EVALUATION

In this section, the Mathematical experimental evaluation of the proposed mathematical model to validate its performanceispresented. The results are obtained using Statistical Analysis Systems (SAS) [46].Table 4 (a) – 4(f) shows results for DS-I, DS-II, DS-III, DS-IV, DS-V, DS-VI respectively. Tables depict the numerical evaluation of the VCM model and show four VCM model output parameters- Actual View Count, Predicted View Count, Neoterics Count, and Followers Count. These Tables shows the predicted view count of various videos' viewers for varying time period. VCM model will keep on improving prediction performance based on system start training and learning of view count parameters. Within 15-45 time stamps system start predicting exactly the same i.e. accurate predicted view count. VCM model another task it to divide Netizens into two groups- Neoterics and Followers, The proposed model is adequately able to depict the actual scenarios. It is also seen from all the tables that the proposed model is adequately able to depict the actual count of two groups' viewers.

Table 4: Model Experimental Results performed for different datasets

Time	View Count	Predicted	Neoterics	Followers
1	14.26	17.85	16.87	0.98
2	51.12	37.34	33.37	3.97
3	77.47	58.49	49.45	9.04
4	96.86	81.33	65.08	16.25
5	114.22	105.83	80.23	25.60
6	122.02	131.94	94.86	37.09
7	201.68	159.58	108.94	50.64
8	203.72	188.61	122.44	66.17
9	265.59	218.87	135.34	83.53
10	267.31	250.15	147.61	102.54
11	305.32	282.22	159.23	122.99
12	306.34	314.80	170.19	144.61
13	340.76	347.63	180.48	167.15
14	344.14	380.42	190.11	190.31
15	371.21	412.86	199.07	213.80
16	407.66	444.69	207.37	237.33
17	460.99	475.66	215.03	260.63
18	471.63	505.53	222.06	283.46
19	524.35	534.11	228.51	305.61
20	536.84	561.25	234.38	326.87
21	577.33	586.84	239.71	347.12
22	603.20	610.79	244.54	366.24
23	622.56	633.06	248.90	384.16
24	670.92	653.66	252.82	400.84
25	705.36	672.59	256.33	416.26
26	735.95	689.91	259.48	430.43
27	789.70	705.68	262.29	443.39

(a) DS-I Musical Dataset

Time	View Count	Predicted	Neoterics	Followers
1	4557.36	2292.10	2276.99	15.11
2	6635.09	4397.20	4341.93	55.27
3	8009.90	6326.00	6212.37	113.63
4	8974.01	8089.30	7904.77	184.53
5	10009.10	9698.30	9434.58	263.72
6	12035.82	11163.7	10816.16	347.54
7	12722.90	12496.2	12062.86	433.34
8	13617.06	13706.0	13186.98	519.02
9	14439.03	14802.9	14199.91	602.99
10	15250.99	15796.3	15112.07	684.23
11	15883.51	16694.8	15933.02	761.78
12	16573.90	17506.8	16671.53	835.27
13	17133.29	18239.9	17335.56	904.34
14	18068.62	18901.2	17932.37	968.83
15	18941.15	19497.3	18468.58	1028.72
16	19613.60	20034.3	18950.17	1084.13
17	20213.16	20517.7	19382.58	1135.12
18	20767.05	20952.7	19770.73	1181.97
19	21361.38	21343.8	20119.06	1224.74
20	21857.52	21695.0	20431.58	1263.92
21	21628.59	22011.5	20711.92	1299.58
22	22228.86	22295.3	20963.36	1331.94
23	22703.98	22550.2	21188.83	1361.37
24	23366.06	22779.0	21390.99	1388.01
25	23816.96	22984.4	21572.22	1412.18
26	24551.56	23168.7	21734.67	1434.03

(b) DS-II Song Dataset

Time	View Count	predicted	Neoterics	Followers
1	446.36	2107	1928.9	177.88
2	4561.55	4099	3487.8	610.64
3	6391.88	5865	4708.0	1156.78
4	8293.40	7345	5636.1	1708.44
5	8782.30	8527	6325.4	2201.18
6	9406.28	9435	6827.5	2607.49
7	9724.55	10113	7187.8	2924.93
8	10123.3	10607	7443.6	3163.76
9	10458.3	10963	7623.7	3338.79
10	10751.7	11215	7749.7	3464.75
11	10877.6	11392	7837.5	3554.31
12	11040.8	11516	7898.6	3617.36
13	11154.2	11602	7940.9	3661.41
14	11308.6	11663	7970.3	3692.19
15	11362.4	11704	7990.6	3713.60
16	11487.9	11733	8004.6	3728.38
17	11559.0	11753	8014.3	3738.68
18	11673.9	11767	8021.0	3745.69
19	11709.8	11776	8025.6	3750.66
20	11788.5	11783	8028.8	3754.07

(c) DS-III Drama Dataset

Time	View Count	Predicted	Neoterics	Followers
1	959.085	1989.75	1963.36	26.39
2	3627.444	3582.45	3498.09	84.36
3	5731.272	4844.09	4691.53	152.56
4	6418.112	5835.24	5615.72	219.52
5	6445.074	6608.83	6329.09	279.74
6	7001.831	7209.55	6878.32	331.23
7	7414.507	7674.19	7300.34	373.85
8	7810.255	8032.48	7624.13	408.35
9	8128.183	8308.11	7872.25	435.86
10	8316.265	8519.77	8062.22	457.55
11	8570.825	8682.07	8207.56	474.51
12	8729.289	8806.4	8318.70	487.70
13	8924.398	8901.56	8403.65	497.91
14	9031.772	8974.34	8468.57	505.77
15	9195.248	9029.99	8518.16	511.83

(d) DS-IV Drama Dataset

Time	View Count	Predicted	Neoterics	Followers
1	9	49	44.34	4.97
2	11	108	87.59	20.82
3	12	178	129.52	48.91
4	266	260	169.92	90.39
5	391	355	208.53	146.01
6	511	461	245.09	215.95
7	683	579	279.38	299.64
8	767	707	311.17	395.62
9	829	842	340.28	501.54
10	949	981	366.60	614.32
11	1200	1120	390.09	730.37
12	1253	1257	410.77	846.00
13	1376	1387	428.75	957.77
14	1428	1507	444.18	1062.81
15	1491	1616	457.27	1159.05
16	1538	1713	468.27	1245.21
17	1708	1798	477.41	1320.83
18	1995	1871	484.96	1386.04
19	2065	1933	491.15	1441.43
20	2101	1984	496.19	1487.91
21	2107	2027	500.28	1526.49
22	2115	2062	503.58	1558.25
23	2125	2090	506.23	1584.20
24	2130	2114	508.36	1605.28
25	2137	2132	510.07	1622.33
26	2142	2147	511.43	1636.06
27	2147	2160	512.52	1647.09
28	2152	2169	513.39	1655.93
29	2156	2177	514.08	1662.99
30	2161	2183	514.63	1668.63
31	2165	2188	515.06	1673.13
32	2169	2192	515.41	1676.71
33	2174	2195	515.69	1679.56
34	2179	2198	515.91	1681.83
35	2185	2200	516.08	1683.63
36	2189	2204	516.48	1687.83
37	2193	2205	516.54	1688.40

(e) DS-V Song Dataset

Time	View Count	Predicted	Neoterics	Followers
1	609.686	582	415.19	166.89
2	1290.521	1101	792.47	308.65
3	1686.817	1563	1132.02	430.97
4	1884.946	1973	1434.74	538.49
5	2405.744	2337	1702.17	634.85
6	2628.657	2659	1936.38	722.78
7	2741.832	2944	2139.83	804.20
8	3002.846	3196	2315.25	880.43
9	3555.527	3418	2465.49	952.25
10	3684.655	3614	2593.38	1020.15
11	3947.72	3786	2701.68	1084.34
12	4017.462	3938	2792.96	1144.92
13	4122.058	4071	2869.59	1201.90
14	4185.888	4189	2933.69	1255.30
15	4292.329	4292	2987.16	1305.10
16	4411.415	4383	3031.66	1351.35
17	4467.221	4463	3068.60	1394.11
18	4499.572	4533	3099.22	1433.46
19	4548.168	4594	3124.56	1469.54
20	4576.699	4648	3145.51	1502.49
21	4625.332	4695	3162.81	1532.48
22	4657.873	4737	3177.08	1559.69
23	4710.039	4773	3188.85	1584.31
24	4741.629	4805	3198.55	1606.52
25	4780.444	4833	3206.54	1626.51
26	4801.868	4858	3213.11	1644.46
27	4841.356	4879	3218.53	1660.55
28	4862.433	4898	3222.98	1674.94
29	4892.746	4914	3226.64	1687.79
30	4904.82	4929	3229.66	1699.26
31	4936.895	4942	3232.13	1709.47
32	4954.024	4953	3234.17	1718.55
33	4986.352	4962	3235.85	1726.62
34	4998.053	4971	3237.22	1733.78
35	5024.404	4978	3238.36	1740.13

(f) DS-IMovie Trailer Dataset

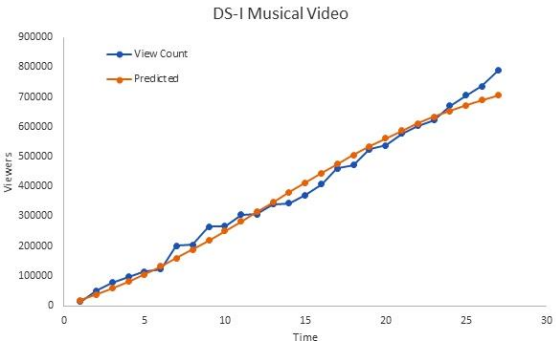
Initially, actual and predicted parameters are validated for the various dataset which is presented in Figure 4(a) - 4(f). It is perceptible in these outcome graphs that the VCM model computed predicted view count of the video is very close to the actual view count.

Video View Count Pattern of actual and predicted results for six videos is shown in Figures 5 (a) -5(f). The graph blue line presents the actual view count of video and the red line denotes the predicted view count. The integrity of the model can be observed based on the goodness of the fit curve for the dataset. It is observed that the predicted and actual values seem to be closely related claiming a fine fitness of curve amongst them. Similar sort of prediction results is generated for all the taken dataset (see Table 2) of the varying category (music, entertainment, comedy, sports, movie, and news & politics) of videos. Hence, evidence of video diffusion patterns can be measured and visualized with the help of the proposed model.

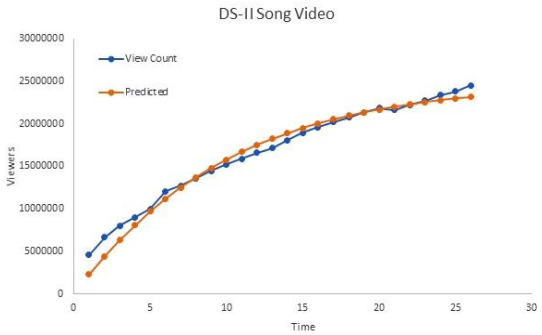
Further, a deep observation about Neoterics and Followers count reflects that in starting Neoterics from start time till timestamp 10, the percentage of Neoterics is higher than the percentage of the followers. But as time increases i.e. at time t=27, the followers are more than the Neoterics. This shows that video is a viral video. It is known that any new video/ product is popular or viral in the market or social media if it is being spread through word of mouth. The comparative analysis of Neoterics and Followers for all six videos is shown in Figure 5(a) - 5(f). These figures depict that DS-I and DS-V are the viral videos while others are not that popular in all the VCM model categorization outcomes pertaining to different videos. In line with what is available in marketing science literature [4, 10, 43, 44].The following observations can be made

- If v1 > v2 then Neoterics takes over the internet marketplace and the maximum level of views is reached at the beginning of the video's life cycle.
- Whereas in reverse case, i.e. when v2 > v1 then followers dominate the market leading to the maximum viewership in the centre of the video's life cycle.

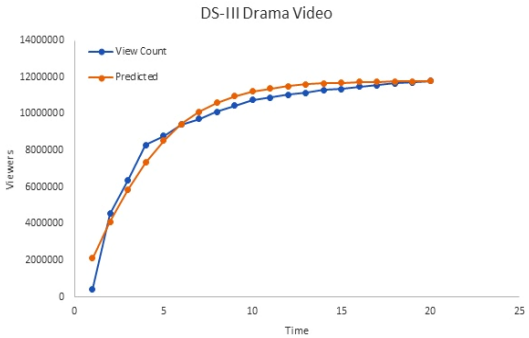
According to the judgment done using the present methodical approach, leaving DS-I, DS-V which have the coefficient of internal influence greater than the coefficient of external influence (i.e. v2 > v1) all the remaining videos have been watched (for the period under consideration) majorly by Neoterics only. Therefore, it is possible to find out the peak for views for only two Datasets here (DS-I and DS-V). We know that there may be many causes behind the virality of a particular online.



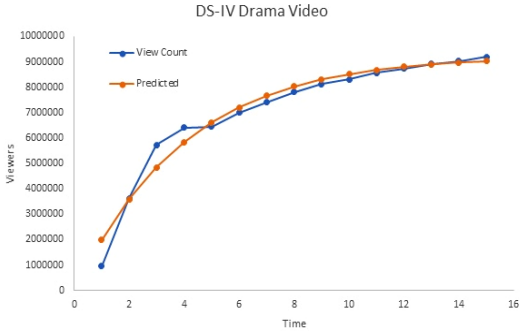
(a) DS-I Musical Video



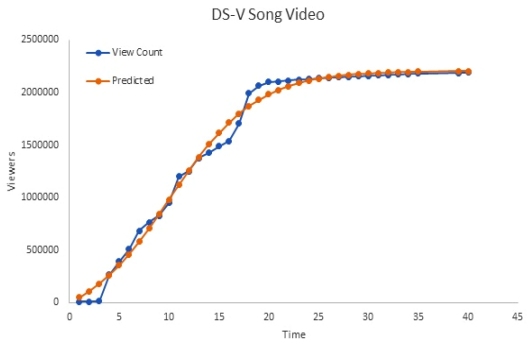
(b) DS-II Song Video



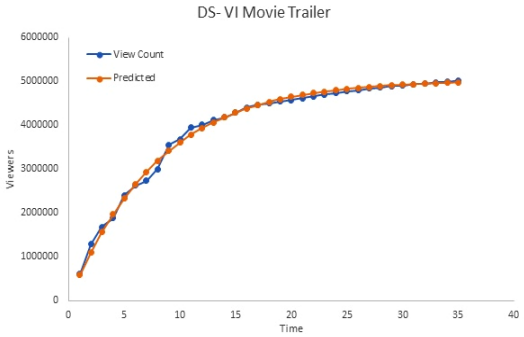
(c) DS-III Drama Video



(d) DS-IV Drama Video



(e) DS-V Song Video



(f) DS-V Movie Trailer Video

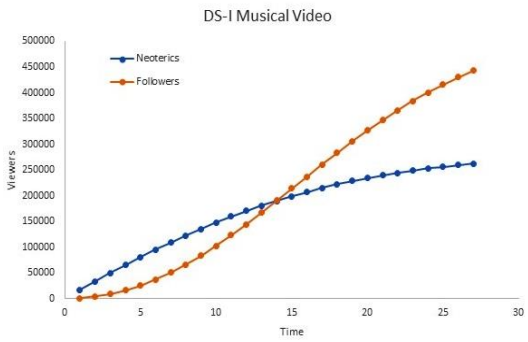
Figure 4(a-f): Actual and Predicted View Counts

5.2. PERFORMANCE MEASURES METRICS:

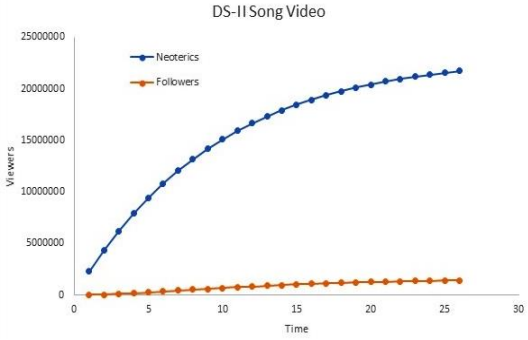
To examine the correctness or accuracy of the model, it is mandatory to measure the fitness of the mathematical model on the real-time datasets and resultant shows that the closeness of the model's outcome towards expectation. Visualization of the fitted plot is given in Figure 4(a-f). Further, it is required to find out the goodness of the proposed model using standard performance metrics [37]. Results are measured using four well-known performance metrics to validate the effectiveness of VCM model; these metrics are Sum of Square Error (SSE), Mean Square Error (MSE), Root Mean Square Error (RMSE),R-Square, and Mean Absolute Percentage Error. These performance evaluation metrics designates that the proposed model has a smaller arbitrary fault component and how good is the prediction capability[4, 5, 24, 31].

- **Sum of Square Error (SSE):** This performance metric evaluates the overall deviation of the predicted view count from the actual view count (see Equation 14).

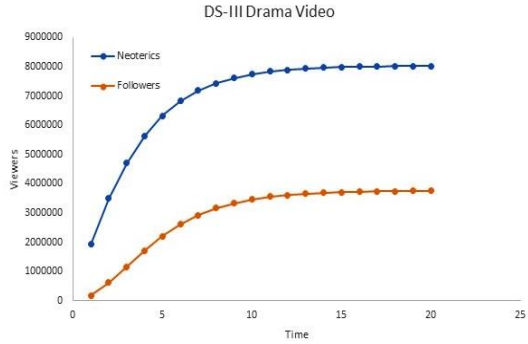
$$SSE = \sum_{time=0}^{t} (Actual\ ViewCount - predicted\ ViewCount)^2 \ldots\ldots (14)$$



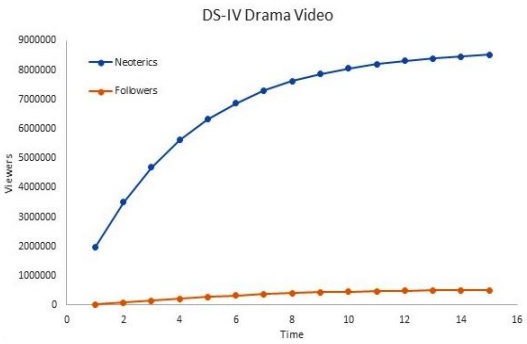
(a) DS-I Musical Video



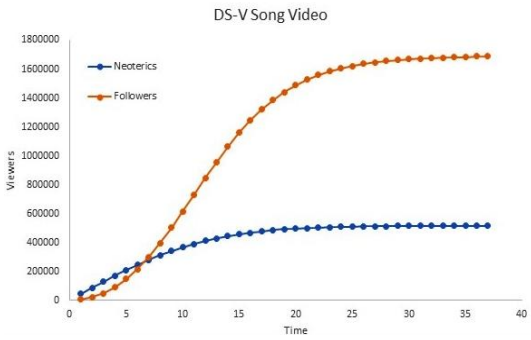
(b) DS-II Song Video

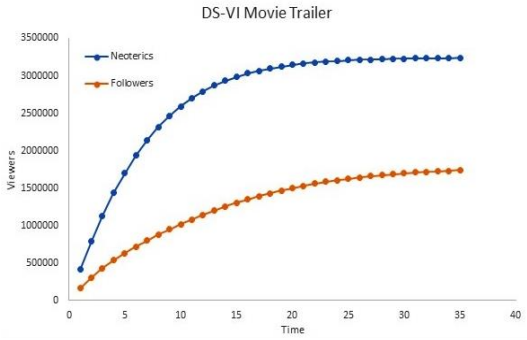


(c) DS-III Drama Video



(d) DS-IV Drama Video





(e) DS-V Song Video (f) DS-V Movie Trailer Video

Figure5(a-f): Neoterics and Follwers View Count Comparison based View Counts

- **Mean Square Error (MSE):** This performance metric is the mean of the overall deviation of the predicted view count from the actual view count (see Equation 15).

$$MSE = \frac{1}{n}\sum_{time=0}^{t}(Actual\ ViewCount - predicted\ ViewCount)^2 \qquad (15)$$

- **Root Mean Square Prediction Error (RMSPE):** It can be obtained using the following equation and represents the Root Mean Square Prediction Error (See Equation 16).

$$RMSPE = \sqrt{(Variance)^2 + (Bias)^2} \qquad (16)$$

Where,

$$Bias = \frac{\sum_{time=1}^{t}(Actual\ ViewCount) - (Predicted\ ViewCount)}{n}$$

And

$$Variance = \sqrt{\frac{\sum_{time=1}^{t}((Actual\ ViewCount - Predicted\ ViewCount)_t - (Bias))^2}{n}}$$

- **R-SQUARE:** It measures the closeness of prediction outcome with the variation of the data(see Equation 17)

$$R - Square = 1 - \left[\frac{\sum_{time=1}^{t}(Actual\ ViewCount - Predicted\ ViewCount)^2}{\sum_{time=1}^{t}\left(Actual\ ViewCount - mean(Actual\ ViewCount)\right)^2}\right] \qquad (17)$$

- **Mean Absolute Percentage Error (MAPE):**It measures the prediction accuracy of a predicting method (see Equation 18).

$$M = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Actual\ ViewCount - predicted\ ViewCount}{Actual\ View\ Count}\right| \qquad (18)$$

All the aforesaid comparison criteria have been evaluated using SAS [46]. After estimation, the parameters of the model and goodness of fit criteria values of six datasets have shown in Table 5. Closure value of R^2 to '1' confirms that our quantified model fits the data reasonably well. Table 5 shows the demonstration video performance evaluation measures.

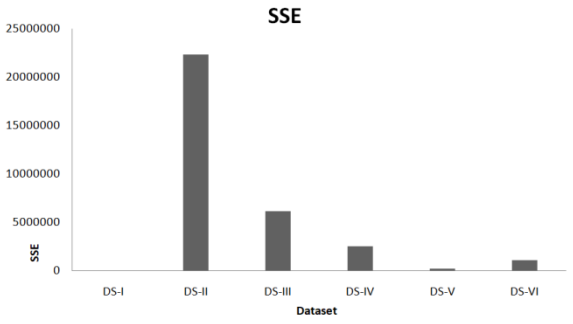
Table 5: Parameter Estimates & Comparison Criteria

	N	v_1	v_2	SSE	MSE	RMSPE	R-Square	MAPE
DS-I	835.52	0.020	0.113	22522.40	8662.46	29.98	0.983	9.2
DS-II	24761.99	0.096	0.014	22314954	929790	964.30	0.973	6.8
DS-III	11797.56	0.180	0.191	6112835	321728	567.55	0.961	22.4
DS-IV	9210	0.240	0.030	2501138	178653	422.70	0.965	10.3

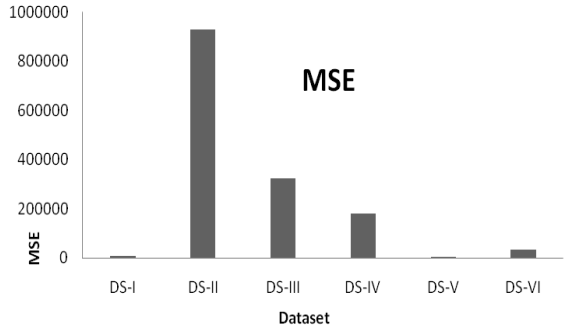
DS-V	2207.34	0.020	0.211	182057.62	4668.14	69184.96	0.991	71.3
DS-VI	5031.40	0.086	0.110	1070000	31460	177.30	0.978	2.3

The distribution of Neoterics coefficient v_1 and Followers coefficients v_2 inferred across the all category of videos is shown in Table 5. Even, Figure 4(a) - 4(f) shows the comfortable and accurate fit of the goodness curve for the proposed model on almost all categories of datasets under consideration and mentioned in Table 2. The datasets results shown in the complete numerical result evaluation section show the average closure value of R^2which is 0.87, 0.91,0.93, 0.86, 0.91 and 0.89 for Music, Entertainment, Comedy, Sports, Movies, and News& Politics videos respectively.

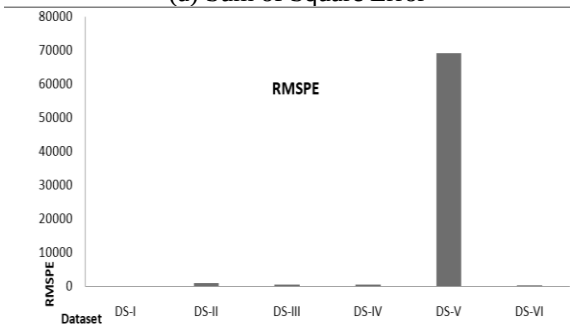
Further, it is available in literature by Lewis et. al. [33] that forecasting is considered to be 'highly accurate' for MAPE values less than 10, 'Good forecasting' for MAPE values ranges between 10-20, 'reasonable' for values ranges between 20-50 and 'inaccurate' in case of values greater than 50. Therefore, it is observed in Figure 6(e) that proposed VCM model is able to achieve good forecasting accuracy for all the datasets except DS-V.A more clear understanding towards all the performance comparison criteria is visualized in Figure 6 (a-e).



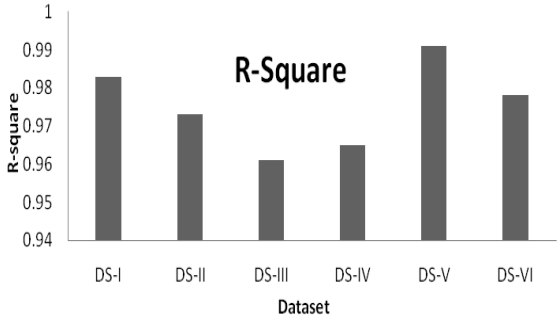
(a) Sum of Square Error



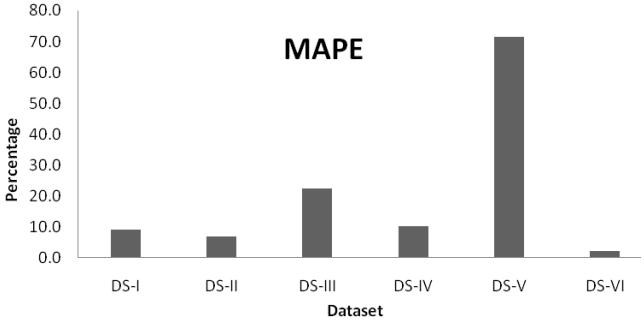
(a) Mean Square Error (MSE)



(c) Root Mean Square Prediction Error (RMSPE)



(d) R-Square



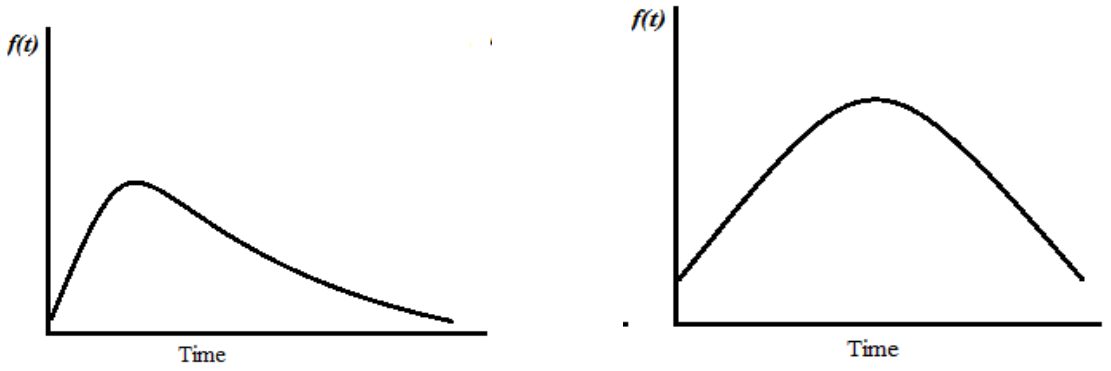
(e) Mean Absolute Percentage Error

Figure 6: Visualization of the evaluation indexes of the model Outcome

6. INTERPRETATION

The important interpretation apparent by the proposed VCM diffusion process model[5] is that Neoterics are initially having much higher importance as compared to followers. Their performance diminishes monotonically with time [31] [13] [25] [19].It shows that the number of viewers through the influence of external forces is more. According to our used numerical interpretation parameter when $v_1 > v_2$ thenNeoterics takes over the internet marketplace and the maximum level of views is reached at the beginning of the video's lifecycle. If the

case is reverse, i.e. when $v_2 > v_1$ then followers dominate the market leading to the maximum sales in the centre of the product's lifecycle. Moreover, when the value of v_1 is lower, then the viewership occurs at a slower rate. Figure 7(a) depicts the viewership growth pattern of internal influence or Neoterics users i.e. v_1 and Figure 7(b) depicts viewership growth pattern of external influence viewers i.e. v_2. Further, it is noticed that for a large value of v_1 and v_2 viewership occurs at a rapid rate and diminish speedily after attaining the maximum level. Thus, various diffusion patterns can be illustrated by altering the values of v_1 and v_2.



(a) $v_2 < v_1$ **(b)** $v_2 > v_1$

Figure 7: Viewership Growth Rate

In the context of online sharing sites, it is not always the case that a person will view the video through WOM (word of mouth) and then activate its diffusion. It might be the case that the video is viewed while surfing the related content online. In such case, v_2 might be or might not always be bigger than v_1 which violates the assumption of standard marketing science literature [10, 35, 43]. This case is an exceptional case that is generally found in the literature of the consumer durable/ service industry. In this scenario, a user watches the video irrespective of his prior knowledge about the offering because of the presence of Netizens/ social media users in today's market. Few videos under this category follow the scenario $v_1 >> v_2$ which means that the current video has not scored in terms of getting popular through word of mouth and the maximum viewership level is reached at the beginning of the life cycle. There might be videos where the counts for a video are because of the viewers who contributed in the viewership through external influence and even down the line the followers do not contribute in the eventual view counts. In this case, the video gains popularity through word of mouth and this case viewership pattern is presented in Figure 7(b) which prevails in the market. Both of these scenarios can be understood in terms of another popularity dynamics named as video virality analysis.

The aforesaid viewership prediction proposal and viewers categorization are crucial for website management which is an important aspect and needs research coverage for the firms. It is a challenging task to know when their offers reach to its maximum virality index.

According to the judgment done using the present mathematical approach shows that DS-I and DS-V videos are getting high virality due to greater internal influence coefficient as compared to external influence coefficient i.e. $v_2 >> v_1$. All other videos are majorly watched byNeoterics. Numerous causes exist behind the virality of a particular online video such as content tagging over social media, people's response to the post, content post time, and attached mentions with content andsocial proofs. These causes provide an added influence to the post/ video. For a specific video in which v_2 does not increase as the time progresses is said to be less influential and lacking trustworthy metrics to enumerate the effectiveness of the posted material. In short, these videos can be termed to become viral through broadcasting. In another case where v_2 takes over v_1, these videos are considered as viral in nature due to word of mouth.

7. CONCLUSION AND FUTURE WORK

This research work discussed and provided a mathematical model. This model works in two directions- 1) It studies the behaviour of YouTube video viewers and predicts the number of YouTube viewers 2) The rate of viewership of a video is classified in two categories: Neoterics: A User Set who are first to view the video and

then activate its diffusion. These users are influenced by an external medium such as advertisements, paid posts. Followers: A User set who view the video through WOM (word of mouth) and then activate its diffusion. These Users are influenced by an internal medium such as trending video, friend's recommendation. The proposed mathematical View Count model fits the real data exceptionally well. This classification plays an important role to understand the virality scenario of the posted videos on YouTube. The study also reveals the point of inflection from where videos gain popularity. The model validation is shown on six different data sets of theYouTube entertainment category. On the other end, model performance is validated on the number of videos lying under various YouTube Categories. Out of all the videos that were taken for validation, the Virality through followers is achieved in two videos- DS-I and DS-Vote concept of these dynamics can be related to the overall viral nature of the particular video which could be due to broadcasting or through word of mouth.

We believe that the work performed in this thesis brings a major contribution towards the aim to analyze and predict the virality of the social media content by providing the content characteristics and statistics based experimental analysis and mathematical modelling based prediction. However, there are few open challenges and opportunities to further improve the virality prediction of the posted content. The present work is focused to predict virality based on view count. It would be interesting and challenging to measure and predict the virality of the posted content based on other content dynamics such as the number of subscribers and other content characteristics such as titles and uploaders of the content. At present, the proposed mathematical model for virality prediction VCM has experimented on YouTube video social platform. This model can be validated for other social media platforms also.

REFERENCES

1. Aggrawal, N., Arora, A., & Anand, A. (2018). Modeling and characterizing viewers of You Tube videos. *International Journal of System Assurance Engineering and Management*, *9*(2), 539-546.
2. Aggrawal, N., Arora, A., Anand, A., & Irshad, M. S. (2018). View-count based modeling for YouTube videos and weighted criteria–based ranking. In *Advanced Mathematical Techniques in Engineering Sciences* (pp. 149-160). CRC Press.
3. Asch, S. E. (1958). A Theory of Cognitive-Dissonance-Festinger, L, 194-195.
4. Anand, A., Agarwal, M., Aggrawal, D., & Singh, O. (2016). Unified approach for modeling innovation adoption and optimal model selection for the diffusion process. *Journal of Advances in Management Research*, *13*(2), 154-178.
5. Anand, A., Agarwal, M., Bansal, G., & Garmabaki, A. H. S. (2016). Studying product diffusion based on market coverage. *Journal of Marketing Analytics*, *4*(4), 135-146.
6. Arora, A., Bansal, S., Kandpal, C., Aswani, R., & Dwivedi, Y. (2019). Measuring social media influencer index-insights from facebook, Twitter and Instagram. Journal of Retailing and Consumer Services, 49, 86-101.
7. Aggrawal, N., Ahluwalia, A., Khurana, P., & Arora, A. (2017). Brand analysis framework for online marketing: ranking web pages and analyzing popularity of brands on social media. Social Network Analysis and Mining, 7(1), 21.
8. Aggrawal, N., Arora, A., & Kumaraguru, P. (2017). Multiple metric aware YouTube tutorial videos virality analysis. International Journal of Social Network Mining, 2(4), 362-387.
9. Bailey, N. T. (1975). *The mathematical theory of infectious diseases and its applications* (No. 2nd edition). Charles Griffin & Company Ltd 5a Crendon Street, High Wycombe, Bucks HP13 6LE.
10. Bass, F. M. (1969). A new product growth for model consumer durables. *Management science*, *15*(5), 215-227.
11. Bisht, M., Irshad, M. S., Aggarwal, N., & Anand, A. (2019, February). Understanding popularity dynamics for YouTube videos: an interpretive structural modelling based approach. In 2019 Amity International Conference on Artificial Intelligence (AICAI) (pp. 588-592). IEEE.
12. Borghol, Y., Mitra, S., Ardon, S., Carlsson, N., Eager, D., & Mahanti, A. (2011). Characterizing and modelling popularity of user-generated videos. Performance Evaluation, 68(11), 1037-1055.
13. Burgess, J., & Green, J. (2018). *YouTube: Online video and participatory culture*. John Wiley & Sons.
14. Cha, M., Kwak, H., Rodriguez, P., Ahn, Y. Y., & Moon, S. (2007, October). 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* (pp. 1-14). ACM.

15. Chakrabarti, D., Wang, Y., Wang, C., Leskovec, J., &Faloutsos, C. (2008). Epidemic thresholds in real networks. *ACM Transactions on Information and System Security (TISSEC)*, *10*(4), 1.
16. Cheng, X., Dale, C., & Liu, J. (2008, June). Statistics and social network of youtube videos. In *2008 16th Interntional Workshop on Quality of Service* (pp. 229-238). IEEE.
17. Chatzopoulou, G., Sheng, C., & Faloutsos, M. (2010, March). A first step towards understanding popularity in YouTube. In 2010 INFOCOM IEEE Conference on Computer Communications Workshops (pp. 1-6). IEEE.
18. Davidson, J., Liebald, B., Liu, J., Nandy, P., Van Vleet, T., Gargi, U., & Sampath, D. (2010, September). The YouTube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 293-296). ACM
19. Dwivedi, Y. K., Williams, M. D., Lal, B., & Schwarz, A. (2008, June). Profiling Adoption, Acceptance and Diffusion Research in the Information Systems Discipline. In ECIS (pp. 1204-1215).
20. Feroz Khan, G., &Vong, S. (2014). Virality over YouTube: an empirical analysis. *Internet research*, *24*(5), 629-647.
21. Granovetter, M. (1978). Threshold models of collective behavior. *American journal of sociology*, *83*(6), 1420-1443.
22. Ganesh, A., Massoulié, L., &Towsley, D. (2005, March). The effect of network topology on the spread of epidemics. In *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.* (Vol. 2, pp. 1455-1466). IEEE.
23. Gürsun, G., Crovella, M., & Matta, I. (2011, April). Describing and forecasting video access patterns. In *2011 Proceedings IEEE INFOCOM* (pp. 16-20). IEEE.
24. Irshad, M. S., Anand, A., & Agarwal, M. (2020). MODELING ACTIVE LIFE SPAN OF YOUTUBE VIDEOS BASED ON CHANGING VIEWERSHIP-RATE. Investigación Operacional, 41(2), 249-262.
25. Irshad, M.S, Anand, A., Bisht, M. (2019). Modelling Popularity Dynamics based on YouTube viewers and Subscribers. International Journal of Mathematical, Engineering and Management Sciences (IJMEMS), Vol. 4(6), 1508-1521.
26. Kapur, P. K., & Garg, R. B. (1992). A software reliability growth model for an error-removal phenomenon. *Software Engineering Journal*, *7*(4), 291-294.
27. Kapur, P. K., Bardhan, A. K., Jha, P. C., & Kapoor, V. K. (2004). An alternative formulation of innovation diffusion model and its extension. *Mathematics and Information Theory*, 17-23.
28. Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, *115*(772), 700-721.
29. Kamiyama, N., & Murata, M. (2019). Reproducing Popularity Distribution of YouTube Videos. IEEE Transactions on Network and Service Management.
30. Kempe, D., Kleinberg, J., &Tardos, É. (2005, July). Influential nodes in a diffusion model for social networks. In *International Colloquium on Automata, Languages, and Programming* (pp. 1127-1138). Springer, Berlin, Heidelberg.
31. Kwon, K. H., & Gruzd, A. (2017). Is offensive commenting contagious online? Examining public vs interpersonal swearing in response to Donald Trump's YouTube campaign videos. Internet Research, 27(4), 991-1010.
32. Lange, P. G. (2007). Publicly private and privately public: Social networking on YouTube. *Journal of computer-mediated communication*, *13*(1), 361-380.
33. Lewis, C. D. (1982). *Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting*. Butterworth-Heinemann.
34. Meyers, L. (2007). Contact network epidemiology: Bond percolation applied to infectious disease prediction and control. *Bulletin of the American Mathematical Society*, *44*(1), 63-86. .
35. Mahajan, V., Muller, E., & Wind, Y. (Eds.). (2000). *New-product diffusion models* (Vol. 11). Springer Science & Business Media.
36. Mir, I. A., & Ur REHMAN, K. (2013). Factors affecting consumer attitudes and intentions toward user-generated product content on YouTube. *Management & Marketing*, *8*(4).

37. Marwell, G., Oliver, P. E., &Prahl, R. (1988). Social networks and collective action: A theory of the critical mass. III. *American Journal of Sociology*, *94*(3), 502-534.
38. Mitzenmacher, M. (2004). A brief history of generative models for power law and lognormal distributions. Internet mathematics, 1(2), 226-251.
39. Paolillo, J. C. (2008, January). Structure and network in the YouTube core. In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008)*(pp. 156-156).
40. Richier, C., Altman, E., Elazouzi, R., Jimenez, T., Linares, G., &Portilla, Y. (2014, August). Bio-inspired models for characterizing YouTube viewcout. In *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)* (pp. 297-305). IEEE.
41. Richier, C., Altman, E., Elazouzi, R., Altman, T., Linares, G., &Portilla, Y. (2014). Modelling view-count dynamics in youtube.
42. Rotman, D., &Preece, J. (2010). The'WeTube'in YouTube–creating an online community through video sharing. *International Journal of Web Based Communities*, *6*(3), 317-333.
43. Rogers, E. M. (1962). Diffusion of InnovationsThe Free Press of Glencoe. *New York*.
44. Ryan, B., & Gross, N. C. (1943). The diffusion of hybrid seed corn in two Iowa communities. *Rural sociology*, *8*(1), 15.
45. Schelling, T. C. (2006). Micromotives and macrobehavior. WW Norton & Company.
46. SAS, (1999) Statistical Analysis Systems. SAS OnLine Doc. Version 8.
47. Susarla, A., Oh, J. H., & Tan, Y. (2012). Social networks and the diffusion of user-generated content: Evidence from YouTube. *Information Systems Research*, *23*(1), 23-41.
48. Singh, O., Anand, A., Kapur, P. K., &Aggrawal, D. (2012). Consumer behaviour-based innovation diffusion modelling using stochastic differential equation incorporating change in adoption rate. *International Journal of Technology Marketing*, *7*(4), 346-360.
49. Szabo, G., & Huberman, B. A. (2008). Predicting the popularity of online content. *Available at SSRN 1295610*.
50. Syed-Abdul, S., Fernandez-Luque, L., Jian, W. S., Li, Y. C., Crain, S., Hsu, M. H., ... & Liou, D. M. (2013). Misleading health-related information promoted through video-based social media: anorexia on YouTube. *Journal of medical Internet research*, *15*(2), e30.
51. Taneja, A., & Arora, A. (2019). Modeling user preferences using neural networks and tensor factorization model. International Journal of Information Management, 45, 132-148.
52. Vaish, A., Krishna, R., Saxena, A., Dharmaprakash, M., & Goel, U. (2012). Quantifying virality of information in online social networks. *International Journal of Virtual Communities and Social Networking (IJVCSN)*, *4*(1), 32-45.
53. Whitler, K. A. (2014). Why word of mouth marketing is the most important social media. *Retrieved January*, *15*, 2018.
54. Williams, M. D., Dwivedi, Y. K., Lal, B., & Schwarz, A. (2009). Contemporary trends and issues in IT adoption and diffusion research. Journal of Information Technology, 24(1), 1-10.
55. Xu, J., Van Der Schaar, M., Liu, J., & Li, H. (2015, April). Timely video popularity forecasting based on social networks. In 2015 IEEE Conference on Computer Communications (INFOCOM) (pp. 2308-2316). IEEE.
56. Zhou, R., Khemmarat, S., & Gao, L. (2010, November). The impact of YouTube recommendation system on video views. In Proceedings of the 10th ACM SIGCOMM conference on Internet measurement (pp. 404-410). ACM.
57. Yang, K. C., Huang, C. H., Yang, C., & Yang, S. Y. (2017). Consumer attitudes toward online video advertisement: YouTube as a platform. Kybernetes.
58. Lebe, S. S., Mulej, M., Batat, W., & Prentovic, S. (2014). Towards viral systems thinking: a cross-cultural study of sustainable tourism ads. Kybernetes.