REVISTA INVESTIGACION OPERACIONAL VOL. 41 , NO. 2, 249-262, 2020

MODELING ACTIVE LIFE SPAN OF YOUTUBE VIDEOS BASED ON CHANGING VIEWERSHIP-RATE

Mohammed Shahid Irshad[*], Adarsh Anand[1*], Mohini Agarwal[**]
[*]Department of Operational Research, University of Delhi, India
[**]Amity School of Business, Amity University, Noida, India

ABSTRACT
Among the various social media sites, YouTube has emerged as a forerunner in the race of online video sharing sites. Being a free uploading and viewing site, YouTube's and the video uploader's main source of revenue are the advertisements run before and during the video. Thus, to better optimize the profits, modeling the view-counts of a particular video is very essential. It is a known fact that the view-count on a video increases with time but the view-count growth rate (viewership rate) is not always increasing. It increases rapidly during the time period when the video becomes viral and it again slows down once this virality phase is over. The behaviour of viewership rate changes multiple times throughout the video life-cycle and leads to altered lifespan of the videos. In the current proposal we have predicted the number of view counts on the basis of changing viewership rate. This change however, might occur several times during video lifecycle, therefore, this ideology has been incorporated in proposed methodical work. A set of models have been discussed which have been ranked using "VIKOR"- multi criterion decision making (MCDM) technique using the YouTube video data sets.

KEYWORDS: Changing Viewership Rate, View-count, VIKOR and YouTube.

MSC: 90B25, 60G99

RESUMEN
Entre los diversos sitios de media-social media, YouTube ha aparecido como el vanguardista en la carrera del intercambio en sitios de video. Siendo un sitio de subida libre, YouTube tiene como la fuente principal de sus ingresos los anuncios que se corren antes y después de descargar el video. Así, lo esencial a optimizar son las ganancias, modelar el número de bajadas de un particular video. Es bien conocido el hecho de qe los conteos de vistas sobre un video se incrementa con el tiempo pero el crecimiento de la tasa de visualizaciones (viewership rate) no siempre es creciente. Esta se incrementa rápidamente durante el periodo de tiempo en que el video se hace viral y nuevamente este decrece lentamente una vez transcurre la fase de la virosis. El comportamiento de la tasa de visualización cambia múltiples veces a través del ciclo de vida del video, lo que conlleva a alterar la amplitud de la vida de los videos. En esta propuesta nosotros hemos predicho el número de conteos de vistas en base a la cambiante tasa de visualización. Este cambio sin embargo, puede ocurrir varias veces durante el ciclo de vida del video, por tanto, esta ideología ha sido incorporada en este trabajo metodológico. Un conjunto de modelos han sido discutidos y ha sido rankeados usando la técnica de toma de decisiones multicriterio (MCDM) "VIKOR"- (MCDM) usando conjunto de datos de YouTube.

PALABRAS CLAVE: tasa cambiante de visualización, , conteo de vistas, VIKOR y YouTube.

1. INTRODUCTION

"On the Origin of Species" was published by English naturalist Charles Darwin in 1859. This book was collection of theories related to the evolution of species on the basis of Charles Darwin's observations. Out of all the theories; theory of natural selection become easiest way to explain the survival and evolution of a specie. One of the basic tenets of this theory was "Individuals who are able to adapt according to the habitat has higher chances of survival". It is true in context of both; mankind as well as technology. There was a time when monopoly of a firm used to be common in many sectors. But with every revolution, either political or industrial; common people got options to choose as per their desire. In political science this power to choose by people later was known as democracy whereas in management science, this led to the competition between firms.

[1] Corresponding Author: adarsh.anand86@gmail.com; mohammedshahid.irshad@gmail.com ;mohini15oct@gmail.com

One such revolution started when a U.S. military funded project “APRANET” came into origin in 1969. This project started the information technology revolution and is now know by what we know as “INTERNET”. In 50 years, internet had grown from connecting five sites to connecting the world. Evolution of internet has not just been limited to the number of device connectivity but also included the enormous applications that we use in our day to day life. With time not only the number of applications increased but also the kind of services provided by the applications has manifolded. Initially the transfer of data was limited to text format but now-a-days data can be transferred in every possible format. The evolution of internet not just increased its reach but also replaced or infiltrated various traditional techniques. With the easy availability of internet and enormous applications we have certainly changed the human various behavioral aspects. Now-a-days everyone is technically hooked to their smart devices and indirectly gathering enormous information from various platforms.

Internet has totally infiltrated the media. In ancient times the announcements were made by humans to inform the population about new laws and regulations. Then written manuscripts came into the picture. After the invention of print technology written manuscripts were replaced by printed pamphlets, books and newspaper (print media) which opened a new gateway for the society to explore new possibilities and made every information accessible to common people. A tremendous change in the human behavior towards news consumption took place when broadcasting media came took over the print media. The emergence of this new medium made a large population to shift themselves from reading newspapers to watching televisions and listening radios.

But like every product, there is some or the other limitation attached to every service or an idea. And when talked in terms of accessibility, “Social Media” has triumphed all the pre-existing mediums as of for now. There are numerous platforms which provide various type of information in several formats across internet. These platforms are known as social media platforms. Even social media platforms have themselves evolved a lot since the inception of first such platform “Sixdegrees.com” in May 1996. Sixdegrees.com was founded by Andrew Weinreich which had features like profile and friend list. Sixdegrees.com was the first social networking website having same format as today’s social media platforms’. It also provided the facility of messaging and bulletin post on others wall depending upon the degree of relationship. Whereas this platform was a closed platform wherein new member could join only through invitation. After this, various social networking platforms were launched but in the world of cut throat competition only those platforms survived who adapted to the user’s requisites. Next platform which became extremely popular was Malaysian built social gaming platform “Friendster” launched in 2002. “LastFm” also made its debut in 2002 which provided online streaming facilities for music and radio programs. Following this, the very next year in 2003, we saw emergence of another giant “LinkedIn” where people connect for business purposes. “Flicker”, “Wordpress”, “SeconfLife”, “Del.icio.us” and “Myspace” and various such data storage and networking sites came into picture in 2003. The year 2004 also has its contribution to social networking. This year saw emergence of sites like those of “Facebook”. Some other sites of that year were like of; “Care2”, “Multiply”, “Ning”, “Orkut”, “Mixi”, “Piczo” and “Hyves”. Initial popularity of Orkut rapidly faded when Facebook swiftly adapted as per user requirement and it technically changed the future of social networking. Right now, Facebook is the one of the most visited websites on internet and also one of the most traffic generating website. YouTube, launched in 2005 changed the world of entertainment. It was the first social network platform which later become the most prominent social media platform. It enabled users to share videos on its platforms. Twitter; launched in 2006 was inspired by the popularity of SMS (Short Message Service): Users could express their thoughts or believes in 140 words on this site. Twitters enabled users to interact with celebrities filling the gap of communication between them. In 2009, another messaging platform “WhatsApp” was launched. It was free, had interactive user interface, supported various files format and did not have word limit. WhatsApp soon became one of the main reason because of which the usage of SMS declined in next few years after its origin. Various other platforms like Instagram (2010), Snapchat (2011), Selfie (2013), Filters (2015), TikTok (2017) were launched in the recent past and have now formally become the life line of human race.

Most of the platforms are free for users and don’t charge for its services. Thus, question arises “what is their source of revenue”. In context of YouTube, it can be said that it is advertisement-based revenue model although YouTube has launched its premium services in 2014. YouTube as well as the content uploader earns through the advertisement. YouTube does not share the earning of channels but we all know revenue generation is directly proportional to view-count of the video [1]. So, it becomes very important to study the behavior of view-count growth so that we can charge the optimal amount for advertisement from advertiser. Richier et al. [16] were one of the first researchers who tried to study the view-count growth pattern. They proposed six different models for viral and non-viral videos. Cheng et al. [8] estimated the active life cycle of videos providing enormous amount of statistics on YouTube. Feroz Khan and Vong [10] studied the impact of other social Medias on generating traffic for YouTube. They stated that most users come to YouTube after finding the link of content on other platforms. Epidemic modelling for understanding the instant view-count was done by Baukhage et al. [6]; wherein he predicted View-count of the video through SIR (Susceptible, Infected, Recovered) modelling in his work. Impact of YouTube recommendation system was studied by the Zhou et al. [21] and Portilla et al.[15]. Bisht et al. [5] found most influential and influenced attributes which derives the view-count of the video by applying ISM (Interpretive Structural Modelling) technique. Aggrawal et al. [2] proposed three models to predict view-count for viral videos. Cheng and Tsai [7] performed sentiment analysis through their proposed framework which is based on deep learning models. Cui et al. [9] also used sentiment analysis for detecting fake news. Yu et al. [18] proposed that a video has multiple phases in its life cycle but there are no models to predict the view-count of the video. Also, they assumed that first phase is always greater than preceding ones.

In the present work, a modelling framework which visualizes and predicts the view-count of the videos incorporating change in the viewership rate aspect has been presented. In line with what has been proposed by various researchers in the past, this phenomenon can be termed as “change point phenomenon”. Change points are the time points after which the rate of viewing of those particular videos increases or decreases. This may happen due to various socio-economic factors like word of mouth, social events etc. There are various videos which are old but came to lime light again because of the recent current affairs causing another active life span for that video. As discussed in literature through epidemic modeling framework, population which is recovered has chances of becoming susceptible (Users thinks about watching that video again) and infected again (Users watched that video again): This also causes changes in the active life span of the lifecycle of videos. Thus, there can be multiple scenarios and many situations when the rate of viewership can get affected. This concept has been utilized in order to study the manifolds that can come in for any video. “Change point” as described in aforesaid lines is not a new ideology and this phenomenon has proved its validity in various and distinct fields like, in software reliability growth modelling change point has been described as “the time where the rate of debugging the product increases or decreases due to change of testing team or learning parameters of the testing team”. [3, 4, 14] In the field of management science also, this fundamental has gained pretty good attention, like that of work by Singh et al. [16], wherein they have proposed a consumer behaviour modeling framework based on change point ideology. A recent study by Irshad et al. [13] discusses about studying the popularity for you tube videos based on Viewers and Subscribers wherein; they have shown that the rate of viewing changes during the lifecycle of video. In the current proposal, in line with Cheng et al. [8] and using the ideology by Irshad et al [13] it has been proposed that the video may encounter changes once or multiple times in its life cycle and thereby the active life span might be altered. As per our available knowledge, this is the first study that discusses such type of framework in context of active life span of a video.

The rest of the paper has been structured as follows: Notations are defined in section 2. Modelling development is discussed in section 3. Section 4 is deals with model validation on real time YouTube video datasets with VIKOR implemented in section 5. Conclusion is presented in section 6 followed by references at the end.

2. NOTATIONS

a	:	Expected number of total Views
b_i	:	Rate of viewing for every interval
$v(t)$	:	Expected number of views by time t
τ_i	:	Time point at which rate of viewing changes
β	:	Scale parameter
k	:	Constant Parameter

3. MODEL DEVELOPMENT

From the work by Aggrawal et al. [2] we consider view-count as a NHPP (Non-Homogenous Poisson process) counting process. Expected number of views in $(t, t+\Delta t)$ is essentially proportional to the expected number of views left from the total expectation at time t, i.e.

$$v(t+\Delta t) - v(t) = b\{a - v(t)\}\Delta t \; + \; o(\Delta t) \tag{1}$$

Dividing equation (1) by Δt and letting $\Delta t \to 0$ following differential equation can be obtained:

$$\frac{dv(t)}{dt} = b(a - v(t)) \tag{2}$$

Solving (2) using initial condition $v(0) = 0$, we have

$$v(t) = a(1 - e^{-bt}) \tag{3}$$

Equation (3) presents a scenario where there is no change throughout the life cycle, we call this to be **Scenario-1**. In line with this, and as part of our proposal, there can be situations when the active life span of the video gets altered at-least once during its life cycle. We call it to be; **Scenario-2**. Like ways, there can be two time periods when this rate gets fluctuated during the life cycle and we term it to be **Scenerio-3.** This can be generalized to study the changes multiple times throughout the life cycle as well.

In line with aforesaid discussion, equation (3) is the expression when the active life span has never been through any fluctuation. On the basis of above explanation, the structure for Scenario-2 can be created using equation (2):

$$\frac{dv(t)}{dt} = b(t)(a - v(t)) \tag{4}$$

Equation (4) represents the view count-based modeling framework for time dependent rate. Since this rate has to be varied at-least once for studying the requisite scenario, the following substitution can be brought out where;

$$b(t) = \left\{ \begin{array}{ll} b_1(t) = b_1 & if \;\; 0 \le t \le \tau \\ b_2(t) = b_2 & if \;\; t > \tau \end{array} \right\} \tag{5}$$

Where τ is the time point when the active life span of the video got changed with respect to the view-counts. Based on the analogy of Huang and Lyu [12] solving the equation (4) by substituting equation (5) at initial condition $v(0) = 0$, we have

$$v(t) = \left\{ \begin{array}{ll} v_1(t) = a(1 - e^{-b_1 t}) & if \;\; 0 \le t \le \tau \\ v_2(t) = a(1 - e^{-b_1\tau - b_2(t-\tau)}) & if \;\; \tau \le t \end{array} \right\} \tag{6}$$

Furthermore, in order to cater to Scenario-3, the following substitution in equation (4) will help us in obtaining the required structure:

$$b(t) = \left\{ \begin{array}{ll} b_1(t) = b_1 & if \;\; 0 \le t \le \tau_1 \\ b_2(t) = b_2 & if \;\; \tau_1 \le t \le \tau_2 \\ b_3(t) = b_3 & if \;\; \tau_2 \le t \end{array} \right\} \tag{7}$$

The mean value function can be obtained by following similar procedure. In this case we have:

$$v(t) = \left\{ \begin{array}{ll} v_1(t) = a(1 - e^{-b_1 t}) & if \;\; 0 \le t \le \tau_1 \\ v_2(t) = a(1 - e^{-b_1\tau_1 - b_2(t-\tau_1)}) & if \;\; \tau_1 \le t \le \tau_2 \\ v_3(t) = a(1 - e^{-b_1\tau_1 - b_2(\tau_2-\tau_1) - b_3(t-\tau_2)}) & if \;\; \tau_2 \le t \end{array} \right\} \tag{8}$$

We can generalize for changes happening n times i.e.

$$b(t)=\begin{cases} b_1(t)=b_1 & if\ \ 0<t\le\tau_1 \\ b_2(t)=b_2 & if\ \ \tau_1<t\le\tau_2 \\ . & . \\ . & . \\ . & . \\ b_n(t)=b_n & if\ \ \tau_{n-1}<t \end{cases} \tag{9}$$

The solution for generalized structure can be obtained in line with [12]:

$$v_n(t)=a\left(1-e^{-(b_n(t-\tau_{n-1})+\sum_{k=1}^{n-1}b_k(\tau_k-\tau_{k-1})}\right) \tag{10}$$

$$where\ \ \tau_0=0$$

Equation (10) technically represents a view count-based model when the rate of viewership before and after time is different but fixed; i.e. it follows an exponential distribution. This distribution is uniformly utilized in studying a phenomenon dealing with constant rate. But in real life, the cases can be a bit different, i.e. the rates before and after the time period under consideration might be dependent on time and need not be fixed. In order to cater to these conceptualizations, we have used certain well-known distributions like that of Weibull, logistic and delayed s shaped.

3.1. Rate as Weibull

Using rate of viewership as Weibull Distribution function and doing the same set of adjustments as done for exponential model in aforesaid scenario, i.e. using Weibull rate in equation (4)

$$b_n(t)=b_nkt^{k-1} \tag{11}$$

In line with [12] The mean value function for n number of changes can be represented as :

$$v_n(t)=a\left(1-e^{-\left(b_n\left(t^k-\tau_{n-1}^k\right)+\sum_{i=1}^{n-1}b_k\left(\tau_i^k-\tau_{i-1}^k\right)\right)}\right) \tag{12}$$

$$where\ \ \tau_0=0$$

3.2. Inflection S-shaped Distribution

Using the rate of viewership as inflexion S shaped Distribution function and performing the similar calculations as done above, the modeling framework to cater to active life span can be derived using the rate as :

$$b_n(t)=\frac{b_n}{1+\beta e^{-b_nt}} \tag{13}$$

Then we get the mean value function can be represented as [12]:

$$v_n(t)=a\left(\frac{1-e^{-\left(b_n(t-\tau_{n-1})+\sum_{k=1}^{n-1}b_k(\tau_k-\tau_{k-1})\right)}}{1+\beta e^{-\left(b_n(t-\tau_{n-1})+\sum_{k=1}^{n-1}b_k(\tau_k-\tau_{k-1})\right)}}\right) \tag{14}$$

$$where\ \ \tau_0=0$$

3.3. Delayed S-shaped Distribution

Following the same procedure as described above, the rate for viewership can be described by

$$b_n(t) = \frac{b_n^2 t}{1+b_n t} \tag{15}$$

Using the mathematical framework by [12], the mean value function can be described as

$$v_n(t) = a\left(1-\left(1+\left(b_n(t-\tau_{n-1})+\sum_{k=1}^{n-1} b_k\left(\tau_k-\tau_{k-1}\right)\right)\right)e^{-\left(b_n(t-\tau_{n-1})+\sum_{k=1}^{n-1} b_k\left(\tau_k-\tau_{k-1}\right)\right)}\right) \tag{16}$$

$where\ \tau_0 = 0$

Table 1: URL's of the YouTube Video

Video I.D.	URL
1	https://www.youtube.com/watch?v=Q6dsRpVyyWs
2	https://www.youtube.com/watch?v=isQ5Ycie73U
3	https://www.youtube.com/watch?v=r6FxROAHJH4
4	https://www.youtube.com/watch?v=oB94lvJbETE
5	https://www.youtube.com/watch?v=GnORm4yR7pg

4. NUMERICAL ILLUSTRATION

In this section, a description of all the different scenarios for every possible case discussed above has been presented. The section contains validation done on five different datasets. The view-count of all five videos was collected after every 24 hours for approximately two months from the date of uploading. The video URL's are provided in Table 1. We had done estimation for three scenarios as discussed on section 3 (Scenario 1, Scenario 2 and Scenario 3) for all distributions using the generalized equation of the distributions i.e. equation 10, 12, 14 and 16. By substituting the value of $n = 1, 2\ and\ 3$ we get the Scenario 1, 2 and 3 for each distribution. The estimated parameters for every distribution are shown in Table 2-5. Parameter estimation is done by using IBM SPSS software.

Table 2: Parameter Estimates when the Distribution is Exponential (using equation (10))

DS 1	**Scenario 1**	**Scenario 2**	**Scenario 3**
a	20287.302	20237.600	20384.816
b	0.293	0.323	0.070
τ		10.000	26.940
b_2		0.010	0.334
b_3			0.156
τ_2			51.658
DS 2	**Scenario 1**	**Scenario 2**	**Scenario 3**
a	3040.204	2981.261	3301.367
b	0.080	0.086	0.012
τ		9.000	29.576
b_2		0.010	0.147
b_3			0.031
τ_2			40.103
DS 3	**Scenario 1**	**Scenario 2**	**Scenario 3**
a	5431.651	5365.072	5491.935

DS 4	**Scenario 1**	**Scenario 2**	**Scenario 3**
a	20403.586	20241.179	20772.495
b	0.133	0.144	0.052
τ		10.000	23.870
b_2		0.010	0.121
b_3			0.043
τ_2			40.214
DS 5	**Scenario 1**	**Scenario 2**	**Scenario 3**
a	1277.714	1264.181	1299.254
b	0.112	0.122	0.043
τ		10.000	22.266
b_2		0.010	0.101
b_3			0.046
τ_2			43.859

b	0.101	0.111	0.019
τ		12.000	29.452
b_2		0.010	0.265
b_3			0.067
τ_2			40.033

Table 3: Parameter Estimates when the distribution is Weibull (using equation (12))

DS 1	**Scenario 1**	**Scenario 2**	**Scenario 3**
a	20576.075	20759.018	20759.018
b	0.445	0.863	0.972
b_2		0.99	0.823
b_3			0.99
τ		40.288	10.034
τ_2			66.873
k	0.683	0.438	0.438
DS 2	**Scenario 1**	**Scenario 2**	**Scenario 3**
a	4386.994	6462.919	4663.693
b	0.173	0.184	0.21
b_2		0.255	0.226
b_3			0.239
τ		18.022	10.31
τ_2			67
k	0.484	0.287	0.397
DS 3	**Scenario 1**	**Scenario 2**	**Scenario 3**
a	5658.374	5866.509	5866.509
b	0.165	0.264	0.316
b_2		0.328	0.276
b_3			0.328
τ		11.275	27.227
τ_2			58.464
k	0.764	0.554	0.554

DS 4	**Scenario 1**	**Scenario 2**	**Scenario 3**
a	21438.253	21261.887	21621.887
b	0.252	0.299	0.314
b_2		0.307	0.269
b_3			0.307
τ		49.751	30.587
τ_2			57.195
k	0.664	0.604	0.604
DS 5	**Scenario 1**	**Scenario 2**	**Scenario 3**
a	1378.383	1630.561	1390.513
b	0.209	0.514	0.714
b_2		0.99	0.01
b_3			0.287
τ		17.697	11.511
τ_2			62.343
k	0.664	0.239	0.585

Table 4: Parameter Estimates when the distribution is Inflection S-shaped Distribution (using (equation (14)))

DS 1	**Scenario 1**	**Scenario 2**	**Scenario 3**
a	20285.918	20384.262	20384.262
b	0.295	0.237	0.081
τ		19.227	19.917
b_2		0.226	0.311

DS 4	**Scenario 1**	**Scenario 2**	**Scenario 3**
a	20399.703	20760.4	20769.984
b	0.134	0.115	0.115
τ		14.192	26.94
b_2		0.097	0.076

b_3			0.226
τ_2			53.579
β	0.01	0.01	0.01
DS 2	**Scenario 1**	**Scenario 2**	**Scenario 3**
a	3038.959	3296.491	3300.354
b	0.081	0.072	0.047
τ		11.385	21.513
b_2		0.044	0.054
b_3			0.043
τ_2			42.529
β	0.01	0.05	0.01
DS 3	**Scenario 1**	**Scenario 2**	**Scenario 3**
a	5430.1	5487.097	5490.932
b	0.102	0.097	0.092
τ		15.055	21.882
b_2		0.089	0.088
b_3			0.087
τ_2			40.693
β	0.01	0.05	0.01

b_3			0.096
τ_2			40.265
β	0.01	0.05	0.01
DS 5	**Scenario 1**	**Scenario 2**	**Scenario 3**
a	1277.332	1298.36	1299.068
b	0.113	0.103	0.069
τ		13.478	22.354
b_2		0.091	0.116
b_3			0.089
τ_2			44.723
β	0.01	0.05	0.01

Table 5: Parameter Estimates when the distribution is Delayed S-shaped(using equation (16))

DS 1	**Scenario 1**	**Scenario 2**	**Scenario 3**
a	20115.493	20354.001	20353.988
b	0.678	0.334	0.638
τ		70.826	10.272
b_2		0.32	0.279
b_3			0.32
τ_2			66.528
DS 2	**Scenario 1**	**Scenario 2**	**Scenario 3**
a	2883.866	3246.686	3246.681
b	0.209	0.091	0.104
τ		42.983	15.827
b_2		0.065	0.082
b_3			0.065
τ_2			45.243
DS 3	**Scenario 1**	**Scenario 2**	**Scenario 3**
a	5271.342	5437.23	5437.228

DS 4	**Scenario 1**	**Scenario 2**	**Scenario 3**
a	19961.822	20636.481	20636.473
b	0.31	0.172	0.086
τ		32.404	10.054
b_2		0.141	0.189
b_3			0.141
τ_2			42.666
DS 5	**Scenario 1**	**Scenario 2**	**Scenario 3**
a	1237.438	1289.959	1289.955
b	0.281	0.17	0.195
τ		22.183	19.196
b_2		0.131	0.122
b_3			0.131
τ_2			62.222

b	0.24	0.148	0.091
τ		53.016	19.262
b_2		0.134	0.197
b_3			0.134
τ_2			44.314

Table 6: Comparison Parameters for All Distribution Models

Comparison Parameters for Exponential Distribution Models															
	DS 1			**DS 2**			**DS 3**			**DS 4**			**DS 5**		
	Scenario 1	**Scenario 2**	**Scenario 3**	**Scenario 1**	**Scenario 2**	**Scenario 3**	**Scenario 1**	**Scenario 2**	**Scenario 3**	**Scenario 1**	**Scenario 2**	**Scenario 3**	**Scenario 1**	**Scenario 2**	**Scenario 3**
R-Square	0.932	0.947	0.947	0.904	0.863	0.971	0.973	0.952	0.979	0.955	0.924	0.986	0.933	0.918	0.942
Variance	648.158	723.174	573.016	200.055	239.491	107.021	185.767	251.157	161.818	786.733	1025.497	435.662	14.266	15.876	11.848
Bias	-20.301	-28.982	0.000	-26.351	-32.552	0.000	-13.032	-27.740	0.000	-82.528	-117.013	0.000	-1.669	-1.857	-1.386
M.S.E.	413037.941	513197.327	323786.601	37374.989	53369.073	11292.272	33513.725	59882.967	25816.258	589545.400	995326.683	187050.244	195.013	241.503	134.509
R.M.P.S.E	648.476	723.755	573.016	201.783	241.693	107.021	186.223	252.684	161.818	791.049	1032.151	435.662	14.364	15.984	11.929
Comparison Parameters for Inflection S-shaped Distribution Models															
	DS 1			**DS 2**			**DS 3**			**DS 4**			**DS 5**		
	Scenario 1	**Scenario 2**	**Scenario 3**	**Scenario 1**	**Scenario 2**	**Scenario 3**	**Scenario 1**	**Scenario 2**	**Scenario 3**	**Scenario 1**	**Scenario 2**	**Scenario 3**	**Scenario 1**	**Scenario 2**	**Scenario 3**
R-Square	0.932	0.947	0.947	0.903	0.970	0.971	0.973	0.979	0.979	0.954	0.985	0.986	0.933	0.941	0.942
Variance	650.368	573.928	573.928	108.030	108.030	107.227	186.786	164.624	162.391	790.909	443.817	437.333	14.311	11.935	11.866
Bias	-20.499	0.025	0.025	0.033	0.033	0.007	-13.176	0.085	0.017	-83.059	0.175	0.035	-1.674	-1.396	-1.388
M.S.E.	415843.821	324817.943	324817.943	11506.169	11506.169	11335.711	33876.712	26719.201	25999.499	595774.618	194118.888	188488.026	196.220	136.484	134.914
R.M.P.S.E	650.691	573.928	573.928	108.030	108.030	107.227	187.250	164.624	162.391	795.258	443.817	437.333	14.408	12.016	11.947
Comparison Parameters for IWeibull Distribution Models															
	DS 1			**DS 2**			**DS 3**			**DS 4**			**DS 5**		
	Scenario 1	**Scenario 2**	**Scenario 3**	**Scenario 1**	**Scenario 2**	**Scenario 3**	**Scenario 1**	**Scenario 2**	**Scenario 3**	**Scenario 1**	**Scenario 2**	**Scenario 3**	**Scenario 1**	**Scenario 2**	**Scenario 3**
R-Square	0.966	0.982	0.982	0.991	0.993	0.992	0.988	0.991	0.991	0.995	0.995	0.995	0.967	0.985	0.974
Variance	457.296	333.578	333.578	60.920	51.580	55.681	123.550	104.338	104.338	261.603	254.778	254.778	6.983	4.607	6.855
Bias	6.831	0.000	0.000	1.114	-0.913	-2.154	3.832	0.000	0.000	2.263	0.000	0.000	-0.817	-0.539	-0.802
M.S.E.	206075.550	109729.066	109729.098	3655.246	2620.549	3042.818	15005.668	10733.004	10733.004	67429.047	63971.238	63971.246	46.719	20.334	45.031
R.M.P.S.E	457.347	333.578	333.578	60.930	51.588	55.723	123.610	104.338	104.338	261.613	254.778	254.778	7.030	4.638	6.902
Comparison Parameters for Delayed S-shaped Distribution Models															
	DS 1			**DS 2**			**DS 3**			**DS 4**			**DS 5**		
	Scenario 1	**Scenario 2**	**Scenario 3**	**Scenario 1**	**Scenario 2**	**Scenario 3**	**Scenario 1**	**Scenario 2**	**Scenario 3**	**Scenario 1**	**Scenario 2**	**Scenario 3**	**Scenario 1**	**Scenario 2**	**Scenario 3**
R-Square	0.845	0.936	0.936	0.765	0.963	0.963	0.910	0.968	0.968	0.854	0.978	0.978	0.864	0.924	0.924
Variance	982.664	628.233	628.233	313.306	120.844	120.844	344.531	200.564	200.564	1423.218	541.731	541.731	19.125	12.789	12.789
Bias	-49.567	2.035	2.049	-41.021	0.859	0.859	-39.595	2.216	2.218	-173.177	3.667	3.669	-2.237	-1.496	-1.496
M.S.E.	944846.811	389182.478	389182.412	91730.230	14395.333	14395.332	112326.190	39644.509	39644.510	1906221.979	289178.851	289178.839	350.439	156.712	156.721
R.M.P.S.E	983.914	628.236	628.236	315.980	120.847	120.847	346.798	200.576	200.576	1433.715	541.743	541.743	19.255	12.876	12.877

Table 6 shows the comparison parameters for each distribution for all datasets. Comparison parameters are calculated to implement VIKOR so we can find the best fitted model out of these.

A notable comparison in the three scenarios can be observed. It can be seen throughout that for a dataset under consideration, Scenario 1, Scenario 2 and Scenario 3 gives a mix type of result i.e. R^2 is better for one scenario and variance is better for another and these results also contradict the hypothesis of Cheng et al. [8] which says videos has single active live span i.e. once the initial burst of the view-count growth completed the video is not going to gather considerable amount of view-count even if it is still on YouTube. From these results we can say that a video can have multiple active phases throughout its life cycle. So, on the well descripted set of graphs (Figure 1-4) have also been presented to showcase the wellness of our proposed models but even then, it is not possible to reach to a conclusion. For overcoming this problem, we have applied a MCDM (Multi Criteria Decision Models) technique name VIKOR which is discussed in section 5.

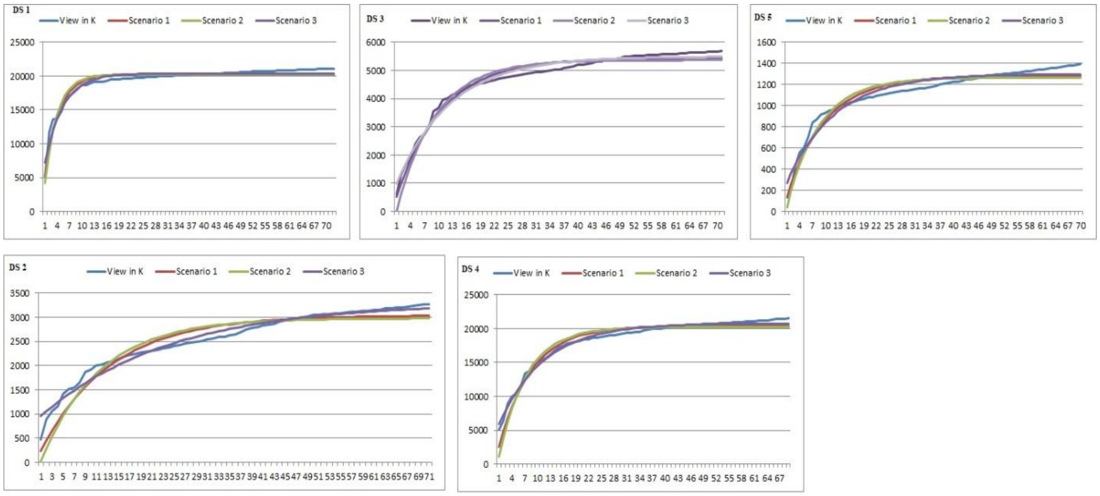

Figure 1: Exponential models for all Datasets

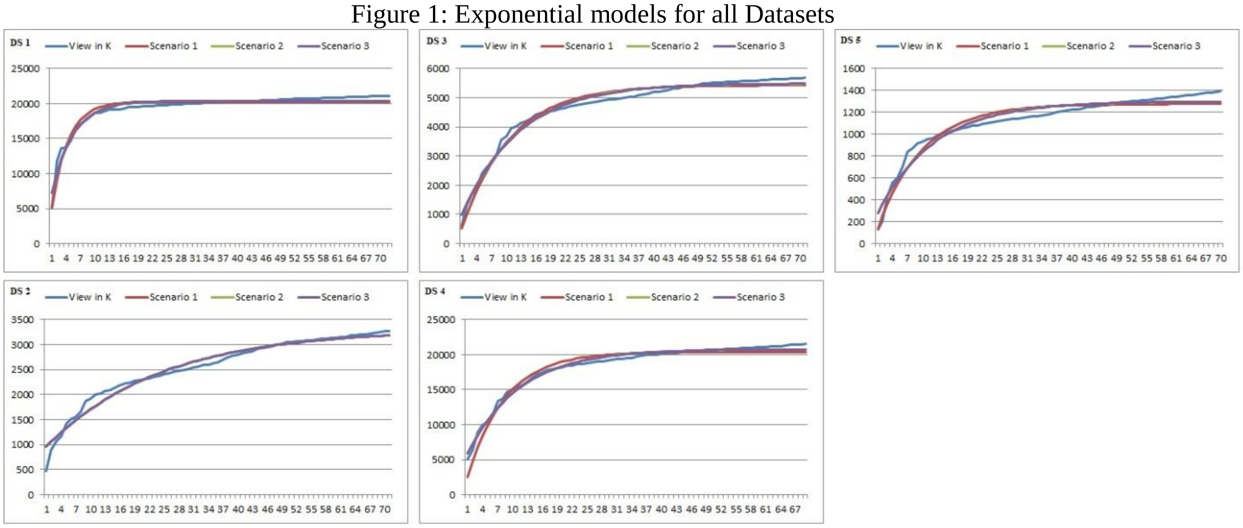

Figure 2: Inflection S-shaped models for all Datasets

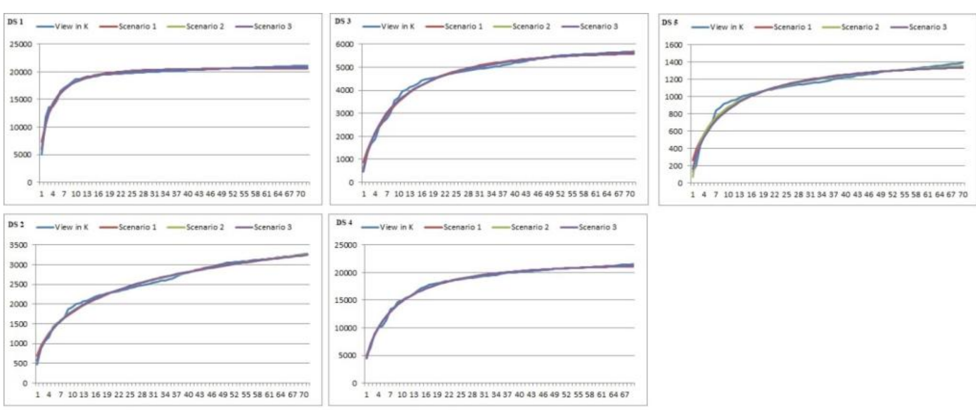

Figure 3: Weibull models for all Datasets

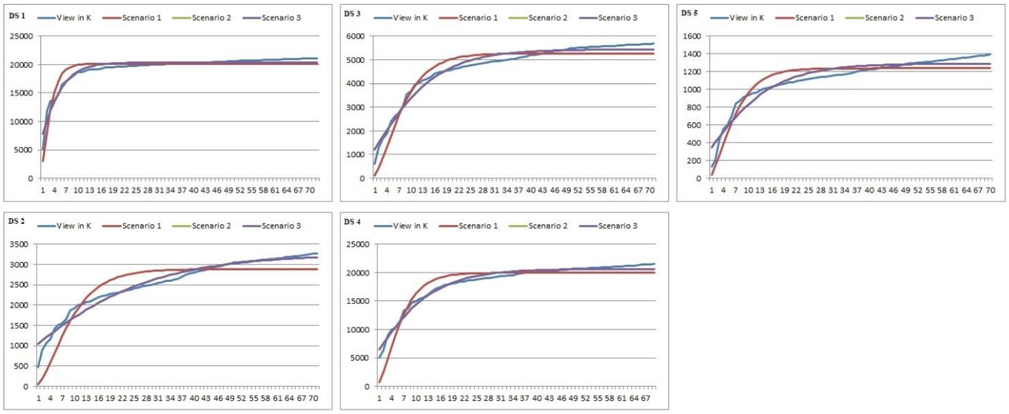

Figure 4: Delayed S-Shaped models for all Datasets

5. VIKOR

The VIKOR method is a multi-criteria decision making (MCDM) or multi-criteria decision analysis method. It was originally developed by Serafim Opricovic to solve decision problems with conflicting and non-commensurable (different units) criteria, assuming that compromise is acceptable for conflict resolution, the decision maker wants a solution that is the closest to the ideal, and the alternatives are evaluated according to all established criteria. VIKOR ranks alternatives and determines the solution named compromise that is the closest to the ideal. The idea of compromise solution was introduced in MCDM by Po-Lung Yu in 1973 [19], and by Milan Zeleny in 1973 [20].

We have to follow seven-step procedure to find the final ranking through VIKOR:

Step-1: Establish a matrix of criteria and different alternatives. (Table 7)

Step-2: Normalization of decision matrix. (Table 8)

Step-3: Calculate the weight of the normalized decision matrix. (Table 8)

$$S_i = \sum_{j=1}^{m}\left(w_j * \left(\frac{x_i^* - x_{ij}}{x_i^* - x_i^-} \right)\right)$$

$$R_i = \max_j \left(w_j * \left(\frac{x_i^* - x_{ij}}{x_i^* - x_i^-} \right)\right) \tag{17}$$

Step-4: Determine the ideal solution and nadir solutions (negative ideal solution): (Table 8)

Ideal Solution | Nadir Solution

$$S^* = \min_i S_i \qquad R^* = \min_i R_i \tag{18}$$

$$S^- = \max_i S_i \qquad R^* = \max_i R_i$$

Step-5: Compute the distance for each alternative.

Step-6: Calculate the relative closeness to the ideal solution.

$$Q_i = v^*\left(\frac{S_i - S^*}{S^- - S^*} \right) + \left(1 - v^*\right)\left(\frac{R_i - R^*}{R^- - R^*} \right) \tag{19}$$

Step 7: Rank the Preference Order. (Table 9)

Table 7: Criteria Matrix with Weights and Best and Worse values

	R-Square	Variance	Bias	M.S.E.	R.M.P.S.E
Scenario 1	0.954	790.909	-83.059	595774.620	795.258
Scenario 2	0.985	443.817	0.175	194118.890	443.817
Scenario 3	0.979	162.391	0.017	25999.499	162.391

Best	0.985	162.391	-83.059	25999.499	162.391
Worst	0.954	790.909	0.175	595774.618	795.258
Weights	0.2	0.2	0.2	0.2	0.2

Table 8: Normalized Matrix with $S_i, R_i, S^*, S^-, R^*, R^-$

	MSE	Bias	Variance	RMSPE	R Sq.	S_i	R_i
Scenario 1	0.200	0.200	0.000	0.200	0.200	0.800	0.200
Scenario 2	0.000	0.090	0.200	0.059	0.089	0.438	0.200
Scenario 3	0.039	0.000	0.200	0.000	0.000	0.238	0.200
					$S^*,\ R^*$	0.238	0.200
					S^-, R^-	0.800	0.200

In this research we have represented VIKOR analysis of DS-4 for Inflection S-shaped distribution models although VIKOR is run on all datasets for each distribution and result is almost same for most of datasets for all distribution models. Equal weight is given to each attribute. From Table 9 we can see that Rank one model is model having two change points. This testifies our hypothesis that a video can have multiple change points in its life-cycle. Table 10 also shows the rank of each scenario for other distributions and from that we can see that change point scenarios have better result than scenario 1.

Table 9: Rank Matrix

Inflection S	S_i	R_i	Q_i	Rank
Scenario 1	0.8	0.2	1	3
Scenario 2	0.437	0.2	0.677	2
Scenario 3	0.238	0.199	0	1

Table 10: Rank Matrix of DS-4 for Exponential, Weibull and Delayed S Models

Exponential	S_i	R_i	Q_i	Rank
Scenario 1	0.497	0.119	0.24728	1
Scenario 2	0.800	0.200	1	3
Scenario 3	0.200	0.200	0.5	2
Delayed S	S_i	R_i	Q_i	Rank
Scenario 1	0.600	0.200	1	2
Scenario 2	0.200	0.200	0	1
Scenario 3	0.200	0.200	0	1
Weibull	S_i	R_i	Q_i	Rank
Scenario 1	0.800	0.200	1	2
Scenario 2	0.000	0.000	0	1
Scenario 3	0.000	0.000	0	1

6. CONCLUSION

The traditional approach of virality largely depends on the total view-count of the video i.e. a video having greater view-count is more viral than a video having comparatively less view-count. But it is not always true higher view-count may be because of broadcasting of content. In this paper we proposed that a video can have multiple change points i.e. the rate of viewing changes multiple times during the lifecycle of video. We estimated different datasets and on different comparison parameters we see the models are performing well especially the models with two change points for all distributions. We can say that the rate of viewing change multiple times due to various external factors like word of mouth or social events. A video having multiple change point will be more viral than no change point. Multiple change points also cause multiple virality phases of a video which can be triggered by various socio-economic factors. Future research can be done for finding multiple viral cycles of in a video life cycle. Advertisers can use the multiple change point for more effective advertising. They will be able to find the time slots when the probability of impact is high.

REFERENCES

[1] AGGRAWAL, N., ARORA, A., and ANAND, A. (2018a): Modeling and characterizing viewers of You Tube videos. **International Journal of System Assurance Engineering and Management,** *9*, 539-546.

[2] AGGRAWAL, N., ARORA, A., ANAND, A., and IRSHAD, M. S. (2018b): View-count based modeling for YouTube videos and weighted criteria–based ranking. In **Advanced Mathematical Techniques in Engineering Sciences**, 149-160 CRC Press, Boca Raton, FL.

[3] ANAND, A., and RAM, M. (Eds.): (2018): **System Reliability Management: Solutions and Technologies**. CRC Press, Boca Raton, FL.

[4] ANAND, A., and RAM, M. (Eds.): (2019**): Recent Advancements in Software Reliability Assurance**. CRC Press, Boca Raton, FL.

[5] BISHT, M., IRSHAD, M. S., AGGARWAL, N., and 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), IEEE**., 588-592, Dubai, UAE.

[6] BAUCKHAGE, C., HADIJI, F., and KERSTING, K. (2015, April): How viral are viral videos? In **Ninth International AAAI Conference on Web and Social Media**, Oxford, London.

[7] CHENG, L. C., and TSAI, S. L. (2019): Deep Learning for Automated Sentiment Analysis of Social Media, Vancouver, Canada.

[8] CHENG, X., DALE, C., and LIU, J. (2008): Statistics and social network of youtube videos. In 2008 **16th International Workshop on Quality of Service, IEEE.**, 229-238, Enschede, Netherlands.

[9] CUI, L., and LEE, S. W. D. (2019): SAME: Sentiment-Aware Multi-Modal Embedding for Detecting Fake News. **The 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining,** Vancouver, Canada

[10] FEROZ KHAN, G., and VONG, S. (2014): Virality over YouTube: an empirical analysis. **Internet Research,** *24*, 629-647.

[11] GOSWAMI, D. N., KHATRI, S. K., and KAPUR, R. (2007): Discrete software reliability growth modeling for errors of different severity incorporating change-point concept. **International Journal of Automation and Computing,** *4*, 396-405.

[12] HUANG, C. Y., and LYU, M. R. (2011): Estimation and analysis of some generalized multiple change-point software reliability models. **IEEE Transactions on Reliability**, *60*, 498-514.

[13] IRSHAD, M.S., ANAND, A. and BISHT, M. (2019): Modelling Popularity Dynamics Based on YouTube Viewers and Subscribers. **International Journal of Mathematical, Engineering and Management Sciences**, *4*, 1508–1521,

[14] KAPUR, P. K., KUMAR, S., and GARG, R. B. (1999): Contributions to hardware and software reliability (Vol. 3): World Scientific, Farrer Road, Singapore.

[15] PORTILLA, Y., REIFFERS, A., ALTMAN, E., and EL-AZOUZI, R. (2015): A Study of YouTube recommendation graph based on measurements and stochastic tools. In **2015 IEEE/ACM 8th International Conference on Utility and Cloud Computing (UCC), IEEE**, 430-435, St. Raphael Resort, Limassol, Cyprus

[16] RICHIER, C., ALTMAN, E., ELAZOUZI, R., ALTMAN, T., LINARES, G., and PORTILLA, Y. (2014): Modelling view-count dynamics in youtube. **arXiv preprint arXiv:1404.2570**.

[17] SINGH, O., ANAND, A., KAPUR, P. K., and AGGRAWAL, D. (2012): Consumer behaviour-based innovation diffusion modelling using stochastic differential equation incorporating change in adoption rate. **International Journal of Technology Marketing**, 7, 346-360.

[18] YU, H., XIE, L., and SANNER, S. (2015): The lifecyle of a youtube video: Phases, content and popularity. In **Ninth International AAAI Conference on Web and Social Media**, Oxford, England.

[19] YU, P. L. (1973): A class of solutions for group decision problems. **Management Science**, 19, 936-946.

[20] ZELENY, M. (1973): **Compromise programming, multiple criteria decision-making. Multiple criteria decision making**. University of South Carolina Press, Columbia.
[21] ZHOU, R., KHEMMARAT, S., and GAO, L. (2010): The impact of YouTube recommendation system on video views. In **Proceedings of the 10th ACM SIGCOMM conference on Internet measureme*nt*, ACM**., 404-410, New York, United States.