# What Determines Viral Phenomenon? Views, Comments and Growth Indicators of TED Talk Videos

Archana Anand Boppolige and Anjula Gurtoo

Abstract-Viral phenomenon drives the business and marketing world to achieve an exponential growth through word-of-mouth diffusion process. The diversity of occurrence of the phenomenon, where a product, online videos, or information, technology, an idea or an event could go viral, pose challenge in measuring which performance measure truly captures viral phenomenon. We performed a measurement study using TED talk videos to identify which of the two performance measures, namely views and comments, truly captures and drives a viral phenomenon. We hypothesized that video comments would be a stronger indicator than video views to measure viral phenomenon. Our results reveal a strong correlation between the two video performance measures. However, video views better represents viral videos and therefore a valid indicator for the measurement of viral phenomenon. The insights can help business strategists and marketing managers in decision making on which performance measure needs business focus in driving virality.

Index Terms—TED talk videos, viral growth measurement, viral marketing, viral phenomenon, Viral videos.

# I. INTRODUCTION

Viral phenomenon, known also as viral marketing [1]-[3] or virality [4]-[5] allures the business and marketing world. It refers to "a word-of-mouth diffusion process wherein a message is actively forwarded from person to person, within and between multiple weakly linked personal networks, and is marked by a period of exponential growth in the number of people who are exposed to the message" [4].

The original inspiration for the term 'viral' in the business world came from the virus like spread of Hotmail, a free web-based e-mail [1], [4], [6], [7]. Hotmail grew subscriptions from zero to 12 million subscriptions in 18 months and at an advertising budget of \$50,000. In comparison, a traditional print publication attains this after a few years of launch [1], [6], [7]. Hotmail is not an isolated case. Another example is the Harry Potter book (Harry Potter and the Deathly Hallows) that set the record for the world's fastest selling book with 15 million copies sold in the first 24 hours [8].

The viral phenomenon spans across different market types and can occur for a product (Example: Blendtec blender) [9], [10], online videos (Example: will.i.am's "Yes We Can" music video) [11], technology (Example: Farmville game on Facebook), information, an idea or an event (2008 U.S. Presidential election) [5]. However, in the present time we are inundated with the term 'viral phenomenon' with no distinction between hype, popularity and what can be termed as a truly viral phenomenon. Moreover, the diversity of the occurrence of viral phenomenon across different medium and market types further adds to the challenge of measuring the viral phenomenon.

In this paper, we attempt to identify a truly viral phenomenon and explore which performance measure can truly capture the phenomenon. Using the data of TED talk videos from the TED website, the objectives of the study are two fold: first, to understand the correlation between various measures of a viral spread and second, to empirically measure the different performance measures on a unifying standard measurement scale and further evaluate the strength of the performance measures in driving a viral phenomenon, and thereby identify a viral phenomenon.

For example, a video can be studied for its virality based on its measures like the number of times a video was viewed, the number of comments the video generated, and also based on the number of times a video was shared with other people. Despite the availability of multiple performance measures, the question remains whether only a particular type of output measure or all performance measures can be valid indicators to measure viral phenomenon?

Two performance measures, namely, views and comments, are taken for TED talk videos. Each performance measure contains three variables each, a variable considering the cumulative total (cumulative comments, cumulative views), and another variable with the highest value of the measure on a single day (peak comments, peak views); the third variable is the exponential growth measured on a daily time variable,

The paper is structured as follows: First, we link our research to the existing literature in viral marketing and detail the motivation for this study. Next, in the methodology section, we describe the data and the measures used in our study, and introduce a novel methodology to empirically measure the different performance measures as a valid indicator to measure viral phenomenon. Next, we present the results of our empirical study. We conclude with discussion on our research findings and opportunities for future research.

# II. LITERATURE REVIEW

Viral phenomenon in the context of business and marketing can be defined as 'a special case of word of mouth diffusion process (can occur either online or offline) [2], [4], [9], [12] with a maximum number of people being exposed and adopting the product [13], [14] within a short duration of time [5], and occurs through personal networks and social influence of people [4], [5], [15].' The analogy of a virus is used to describe

Manuscript received November 28, 2016; revised January 6, 2017.

Archana Anand Boppolige is a Research Scholar with the Department of Management Studies, Indian Institute of Science, Bangalore, 560012 India (e-mail: baarchana@mgmt.iisc.ernet.in).

Anjula Gurtoo is an Associate Professor with the Department of Management Studies, Indian Institute of Science, Bangalore, 560012 India (e-mail: anjula@mgmt.iisc.ernet.in).

the diffusion of information or a product [14], where the viral spread occurs exponentially from one consumer to the next [1], [2], [4].

The studies on viral phenomenon growth measurement and in particular the research on multiple performance measures and its validity to measure a viral phenomenon are scant [12], [16]. Previous researchers have used measures such as emails (to email forwarding behavior), mass test media advertisements (television, newspaper), and website hits such as YouTube video views, twitter tweets and blog postings. The studies using these measures range from content analysis, social ties and networks, and the type of sharing mechanism that drives a viral phenomenon. There are also studies that understand the propagation pattern of the measures, mostly for videos, from a diffusion point of view and looks at the popularity of videos without dealing with its viral growth.

Two studies are of interest to us because of their measurement driven analysis [17], [18]. Reference [17] studied the propagation pattern of YouTube videos using the measure of video views to understand how the YouTube recommendation system on video views impacts the popularity of the videos. The findings from the study show a strong correlation exits between the measures YouTube video views and videos comments. Although the study was conducted using a single measure of video views but recommends using comments as another measure to predict video popularity. Reference [18] used the measure of favorite count (pictures marked as favorite by users) to study the evolution of popularity of pictures over time on Flickr, a photo sharing social network. The study, used a single measure of favorite count to study popularity and also checks for the relationship between number of picture views, comments and fans. The findings show a low correlation between views and comments but a high correlation between number of fans and number of comments.

The set of studies that have considered multiple measures to study viral phenomenon predominantly explore the perspective of sharing mechanism. [19] studied YouTube videos using the measure of video views to analyze the sharing mechanism (internal YouTube search and external links) and the relationship to video popularity. References [5] and [11] used the measure of YouTube video views to study the relationship between video viewership and the blogs linking to the videos. The focus of the studies has been on understanding the interplay of the measures and the type of sharing mechanism that drives virality. However, the validity of the measures per say to test if the performance measures considered is a valid indicator to measure viral phenomenon has been vastly ignored.

Existing literature recommends using measures other than views, like comments, to understand virality. With no prior studies conducted, there exists a gap in the empirical validity of the different performance measures that can be used to measure viral phenomenon i.e., to validate which performance measure is a valid indicator to measure viral phenomenon.

# III. METHODOLOGY

#### A. Data Description

The data set consists of TED talk videos posted on the TED

website, www.ted.com, between 21 November, 2013 and 18 January, 2014. There were 34 TED talk videos posted during this period. The information for each talk -Talk title, speaker name, number of views and number of comments- are tracked daily for 46 days for comments and extended to 134 days for video views since the online posting of the video using a python script. TED talk videos are selected as our data for analysis because all videos are free to watch online with no influence of favorable advertisement towards any particular TED video.

# B. Measures

The scope of our analysis is two TED talk video measures namely views and comments. Views represent the number of times a video was viewed. The TED video on the website also allows qualitative feedback through the viewers' comments.

We argue that commenting on a video or sharing a video represents a much deeper engagement and interaction, and is a larger investment of time and effort than watching a video. In contrast, views can be confounded by seeding tactics, difficult to separate individual occurrences, and there could be instance when a video is playing in the background adding to the views without the person really paying attention. Owning to the above factors, we hypothesize that video comments would be a stronger indicator to measure viral phenomenon. Therefore, in this analysis, we focus on the number of comments to measure the growth of viral phenomenon and compare by validating using views as a viral phenomenon growth measure.

The variables used in the analysis under the measure 'comments' include cumulative number of comments (for 46 days after online video posting), maximum number of comments on any single day during the 46 days time period termed as peak comments, and the time variable in terms of days measured for exponential growth.

The variables for the measure 'video views', are cumulative number of views (for 134 days after online video posting) and maximum number of views on any single day during the 134 day time period termed as peak views. The performance measure of views on a time variable of days is also measured for exponential growth. It must be noted that the peak comments and peak views is a maximal from the daily number of comments and views during a defined time period.

To summarize, we considered two performance measures namely, comments and views, for the analysis of TED talk videos, with each measure containing three variables, namely cumulative, peak, and exponential growth.

#### C. Representativeness of Tracked Data

To verify representativeness of the tracked views and comments, the percentage of the tracked views and comments (for 46 days since the online video posting) are calculated from the total views and comments of a video at the end of two years after the online posting of the video i.e., as on 30 December 2015. Here, tracked comments and views represents the time period during which the information of TED talk videos was recorded. The total comments and total views represents the total number of comments and views at the end of two years after the online posting of the video i.e., as on 30 December 2015. For instance, Talk ID 1560 contains 228 comments on 46<sup>th</sup> day since the online posting of the video. The number of

comments for the talk after nearly two years of online posting of the video i.e., as on 30 December 2015, is 242 comments.

Table 1 gives the results. The talk shows 94.21 percentage representativeness of the tracked data which means 46 days of tracking covers most of the comments (94.21%) for the talk. The result in Table I show that a majority of TED talk videos (32 TED talk videos or 94% of videos in our data set) shows greater than 70 percentage representativeness of the tracked comments for 46 days after the online posting of the video. Whereas the representation of tracked views for 46 days after the online posting of the video are not highly representative with only one video showing greater than 70 percentage representativeness. Therefore, we extend the tracking of daily video views for 134 days of viewing days. With the extended time period, the tracked views show a high representation of the total views with 26 TED talk videos showing higher than 70 percentage representativeness with tracked data.

The result lead to our conclusion that the tracked comments for 46 days and the tracked views for 134 days after the online posting of the video represents the total comments and total views very well. This also means video views have a wider spread and sustains for a longer time when compared to video comments. This is also supported by the wide variation in the days for attaining the maximum views and comments.

| Represe   | Representativeness   | Representativeness of views |                      |  |  |  |
|-----------|----------------------|-----------------------------|----------------------|--|--|--|
| ntativene | of comments          |                             |                      |  |  |  |
| ss (%)    | For 46 days          | For 46 days                 | For 134 days         |  |  |  |
| Below     |                      | Talk IDs 1552, 1554,        |                      |  |  |  |
| 30%       |                      | 1555, 1556, 1557,           |                      |  |  |  |
|           | -                    | 1558, 1559, 1562,           | -                    |  |  |  |
|           |                      | 1564, 1568, 1573            |                      |  |  |  |
|           |                      | [Count =11 videos]          |                      |  |  |  |
| 30 to     |                      | Talk IDs 1550, 1551,        | Talk ID 1568         |  |  |  |
| 49%       |                      | 1553, 1560, 1561,           | [Count =1 video]     |  |  |  |
|           | -                    | 1563, 1565, 1570,           |                      |  |  |  |
|           |                      | 1576                        |                      |  |  |  |
|           |                      | [Count =9 videos]           |                      |  |  |  |
| 50 to     | Talk IDs 1554, 1568  | Talk IDs 1566, 1567,        | Talk IDs 1554, 1555, |  |  |  |
| 69 %      | [Count =2 videos]    | 1569, 1571, 1572,           | 1570, 1571, 1573,    |  |  |  |
|           |                      | 1574, 1575, 1577,           | 1576, 1579           |  |  |  |
|           |                      | 1578, 1579, 1580,           | [Count =7 video]     |  |  |  |
|           |                      | 1581, 1582                  |                      |  |  |  |
|           |                      | [Count =13 videos]          |                      |  |  |  |
| 70 to     | Talk IDs 1555, 1558, | Talk ID 1583                | Talk IDs 1550, 1551, |  |  |  |
| 90 %      | 1562, 1564, 1566,    | [Count =1 video]            | 1552, 1553, 1556,    |  |  |  |
|           | 1570, 1571, 1572,    |                             | 1557, 1558, 1559,    |  |  |  |
|           | 1573, 1574, 1576,    |                             | 1560, 1561, 1562,    |  |  |  |
|           | 1577, 1578, 1580,    |                             | 1563, 1564, 1565,    |  |  |  |
|           | 1581, 1582           |                             | 1566, 1567, 1569,    |  |  |  |
|           | [Count =16 videos]   |                             | 1572, 1574, 1575,    |  |  |  |
|           |                      |                             | 1577, 1578, 1580,    |  |  |  |
|           |                      |                             | 1581, 1582, 1583     |  |  |  |
|           |                      |                             | [Count =26 videos]   |  |  |  |
| Above     | Talk IDs 1550, 1551, |                             |                      |  |  |  |
| 90%       | 1552, 1553, 1556,    |                             |                      |  |  |  |
|           | 1557, 1559, 1560,    |                             |                      |  |  |  |
|           | 1561, 1563, 1565,    | -                           | -                    |  |  |  |
|           | 1567, 1569, 1575,    |                             |                      |  |  |  |
|           | 1579, 1583           |                             |                      |  |  |  |
|           | [Count =16 videos]   |                             |                      |  |  |  |

# D. Methodology for Validating the Multiple Measures

The theory on a viral phenomenon following exponential growth is strongly acknowledged and established in viral marketing despite the limited empirical studies [1]-[2], [4]. Therefore, we identify a viral video, irrespective of the measure of comments and views, by basing it on the exponential growth model. A video is considered viral if its follows an exponential growth pattern.

The two main aspects of a viral spread are speed and reach [4]. The speed is captured by the exponential growth component and the reach signifies how widely the phenomenon propagates. Through our methodology, we capture both aspects of viral phenomenon while also validating the suitability of the different performance measures to measure viral phenomenon. First the videos are empirically identified for high performing videos using a common measurement score across the different measures and further checked for viral growth using the curve fitting method. The methodology is tested in the results section.

#### IV. RESULTS

#### A. Correlation between Comments and Views

| TABLE II: CHECK FOR | NORMALITY FOR COMMENTS |
|---------------------|------------------------|
|---------------------|------------------------|

| TABLE II: CHECK FOR NORMALITY FOR COMMENTS |      |           |                      |                    |  |  |
|--|------|-----------|----------------------|--------------------|--|--|
| Talk Skew Kurtosi                          |      | Kurtosis  | Kolmogorov-          | Shaniro-Wilk       |  |  |
| ID   | ness | Kui tosis | Smirnov <sup>a</sup> | Shaph 0- Wilk      |  |  |
| 1550                                       | 3.14 | 11.98     | D(44)=0.29, p< .001  | W(44)=0.60, p<.001 |  |  |
| 1551                                       | 3.14 | 12.75     | D(44)=0.29, p< .001  | W(44)=0.61, p<.001 |  |  |
| 1552                                       | 2.97 | 11.06     | D(44)=0.28, p< .001  | W(44)=0.62, p<.001 |  |  |
| 1553                                       | 2.35 | 6.02      | D(44)=0.29, p< .001  | W(44)=0.63, p<.001 |  |  |
| 1554                                       | 2.97 | 9.62      | D(44)=0.29, p< .001  | W(44)=0.58, p<.001 |  |  |
| 1555                                       | 1.84 | 2.34      | D(44)=0.32, p< .001  | W(44)=0.66, p<.001 |  |  |
| 1556                                       | 2.29 | 3.98      | D(44)=0.39, p<.001   | W(44)=0.52, p<.001 |  |  |
| 1557                                       | 2.49 | 5.50      | D(44)=0.37, p<.001   | W(44)=0.54, p<.001 |  |  |
| 1558                                       | 3.20 | 11.17     | D(44)=0.34, p<.001   | W(44)=0.51, p<.001 |  |  |
| 1559                                       | 2.21 | 3.49      | D(44)=0.41, p<.001   | W(44)=0.51, p<.001 |  |  |
| 1560                                       | 2.69 | 6.91      | D(44)=0.39, p<.001   | W(44)=0.52, p<.001 |  |  |
| 1561                                       | 2.69 | 6.41      | D(44)=0.35, p<.001   | W(44)=0.52, p<.001 |  |  |
| 1562                                       | 5.05 | 27.14     | D(44)=0.41, p<.001   | W(44)=0.32, p<.001 |  |  |
| 1563                                       | 6.08 | 38.87     | D(44)=0.39, p<.001   | W(44)=0.28, p<.001 |  |  |
| 1564                                       | 1.64 | 2.57      | D(44)=0.21, p<.001   | W(44)=0.81, p<.001 |  |  |
| 1565                                       | 3.27 | 11.03     | D(46)=0.36, p<.001   | W(46)=0.50, p<.001 |  |  |
| 1566                                       | 5.12 | 29.74     | D(44)=0.34, p<.001   | W(44)=0.41, p<.001 |  |  |
| 1567                                       | 5.03 | 29.19     | D(43)=0.33, p<.001   | W(43)=0.43, p<.001 |  |  |
| 1568                                       | 4.29 | 21.52     | D(43)=0.28, p< .001  | W(43)=0.52, p<.001 |  |  |
| 1569                                       | 4.32 | 21.99     | D(43)=0.32, p<.001   | W(43)=0.49, p<.001 |  |  |
| 1570                                       | 3.67 | 15.42     | D(43)=0.30, p<.001   | W(43)=0.54, p<.001 |  |  |
| 1571                                       | 2.02 | 3.86      | D(43)=0.25, p<.001   | W(43)=0.72, p<.001 |  |  |
| 1572                                       | 5.59 | 33.77     | D(43)=0.41, p<.001   | W(43)=0.31, p<.001 |  |  |
| 1573                                       | 4.17 | 19.03     | D(43)=0.34, p< .001  | W(43)=0.44, p<.001 |  |  |
| 1574                                       | 3.51 | 12.37     | D(43)=0.39, p<.001   | W(43)=0.43, p<.001 |  |  |
| 1575                                       | 3.85 | 15.86     | D(43)=0.35, p<.001   | W(43)=0.44, p<.001 |  |  |
| 1576                                       | 3.09 | 9.55      | D(43)=0.36, p<.001   | W(43)=0.50, p<.001 |  |  |
| 1577                                       | 3.99 | 19.07     | D(43)=0.36, p<.001   | W(43)=0.47, p<.001 |  |  |
| 1578                                       | 3.98 | 17.57     | D(43)=0.37, p<.001   | W(43)=0.45, p<.001 |  |  |
| 1579                                       | 4.39 | 22.32     | D(43)=0.35, p<.001   | W(43)=0.43, p<.001 |  |  |
| 1580                                       | 4.48 | 22.54     | D(43)=0.37, p<.001   | W(43)=0.39, p<.001 |  |  |
| 1581                                       | 4.98 | 27.77     | D(43)=0.37, p<.001   | W(43)=0.37, p<.001 |  |  |
| 1582                                       | 3.52 | 12.43     | D(43)=0.39, p<.001   | W(43)=0.43, p<.001 |  |  |
| 1583                                       | 3.05 | 9.61      | D(43)=0.39, p<.001   | W(43)=0.48, p<.001 |  |  |

a. Lilliefors Significance Correction

We first check if there exists a statistically significant relationship between number of video comments and video

views. A bi-variate correlation analysis, Spearman's rho correlation, is performed to find the association between the TED video comments and TED video views.

Spearman's rho correlation analysis is performed because the assumptions of normality, linearity and presence of no outliers are violated for all the 34 TED talk videos for the variable comments. A similar check for the variable views also results in scores that are significantly different from a normal distribution when tested for both 46 days and 134 days of viewing. The results in Table II show a high positively skewed distribution for comments for all the 34 TED talk videos. The kurtosis shows a leptokurtic distribution i.e., a heavy-tailed positive kurtosis which indicates a deviation from normal. The test statistic for both Kolmogorov-Smirnov (K-S) and Shapiro-Wilk shows significantly non-normal at 0.001.

| Г | AB | LE | Π | I: | CORRELATION I | BETWEEN | COMMENTS AND | VIEWS |
|---|----|----|---|----|---------------|---------|--------------|-------|
|   |    |    |   |    |               |         |              |       |

| Strength of the<br>relationship | Spearman's correlation coefficient (r <sub>s</sub> )   |
|---------------------------------|--|
| No relationship (0)             | -  |
| Small effect (±0.1)             | -  |
| Medium effect<br>(±0.3)         | Talk ID 1583 ( $r_s = 0.43^{**}$ ), Talk ID 1582 ( $r_s = 0.49^{**}$ ),<br>Talk ID 1579 ( $r_s = 0.42^{**}$ ), Talk ID 1577 ( $r_s = 0.47^{**}$ ),<br>Talk ID 1572 ( $r_s = 0.47^{**}$ ), Talk ID 1561 ( $r_s = 0.49^{**}$ )<br>[Count = 6 videos]   |
| Large effect (±0.5)             | $ \begin{array}{l} \mbox{Talk ID 1581 (r_s = 0.63^{**}), \mbox{Talk ID 1580 (r_s = 0.58^{**}), \mbox{Talk ID 1578 (r_s = 0.63^{**}), \mbox{Talk ID 1576 (r_s = 0.62^{**}), \mbox{Talk ID 1577 (r_s = 0.63^{**}), \mbox{Talk ID 1577 (r_s = 0.63^{**}), \mbox{Talk ID 1579 (r_s = 0.63^{**}), \mbox{Talk ID 1569 (r_s = 0.71^{**}), \mbox{Talk ID 1568 (r_s = 0.62^{**}), \mbox{Talk ID 1567 (r_s = 0.67^{**}), \mbox{Talk ID 1566 (r_s = 0.53^{**}), \mbox{Talk ID 1566 (r_s = 0.53^{**}), \mbox{Talk ID 1566 (r_s = 0.53^{**}), \mbox{Talk ID 1566 (r_s = 0.57^{**}), \mbox{Talk ID 1562 (r_s = 0.50^{**}), \mbox{Talk ID 1560 (r_s = 0.61^{**}), \mbox{Talk ID 1559 (r_s = 0.54^{**}), \mbox{Talk ID 1556 (r_s = 0.53^{**}), \mbox{Talk ID 1560 (r_s = 0.53^{**}), Talk ID 1560 (r_s $ |
|                                 | [Count = 16 videos]  |
| Very strong effect<br>(±0.75)   | Talk ID 1574 ( $r_s = 0.84^{**}$ ), Talk ID 1573 ( $r_s = 0.75^{**}$ ),<br>Talk ID 1570 ( $r_s = 0.88^{**}$ ), Talk ID 1564 ( $r_s = 0.84^{**}$ ),<br>Talk ID 1558 ( $r_s = 0.73^{**}$ ), Talk ID 1557 ( $r_s = 0.73^{**}$ ),<br>Talk ID 1555 ( $r_s = 0.79^{**}$ ), Talk ID 1554 ( $r_s = 0.84^{**}$ ),<br>Talk ID 1553 ( $r_s = 0.80^{**}$ ), Talk ID 1552 ( $r_s = 0.83^{**}$ ),<br>Talk ID 1551 ( $r_s = 0.84^{**}$ ), Talk ID 1550 ( $r_s = 0.76^{**}$ )<br>[Count = 12 videos]   |
| Perfect (+1)                    | -<br>-   |

\*\*Correlation is significant at the 0.01 level (2-tailed)

The results of Spearman's correlation analysis shows a significant (p < 0.01, for two-tailed) and positive correlation between TED video comments and TED video views for the entire 34 TED talk videos. The Spearman's correlation coefficient  $(r_s)$  and the strength of the relationship are shown in Table III. We interpret the strength of the relationship as given by Cohen [20] where r = 0.10, 0.30 or 0.50 represents small, medium and large effect respectively. The results show that 82% (28 of the 34 videos exhibit a large correlation effect where effect size is greater than 0.50. Therefore, we further fine tune the effect size on the extreme to define a higher strength of r = 0.75 to reflect very strong effect. From Table III, we summarize the strength of the correlation relationship with 6, 16 and 12 TED talk videos exhibiting medium effect, large effect and very strong correlation effects respectively between the two variables, comments and views.

#### B. Identifying High Performing Videos

A video is considered as a high performing video if it

contains the maximum number of comments in comparison to other videos in the data set. The maximum number of comments could be either the maximum comments on any single day (peak comments) or at the end of a defined time period (for instance 46 days after online video posting) at the cumulative level. The measure views is also considered in our viral phenomenon measurement analysis. We adopted two different methods (Method 1: using z-scores, Method 2: using box plot) to identify the outliers of high performing videos for two different measures, namely comments and views.

Identifying outliers (high performing videos) using z-scores: In this section we detail the method of identifying outliers of high performing videos using the method of z-scores. Since there are two performance measures considered for the videos, comments and views, with each measure containing two variables, namely cumulative comments, peak comments, cumulative views and peak views respectively, therefore, the z-score method helps to standardize the scores across the different measures and its corresponding variables. The actual values are converted into their respective z-scores for all the variables and the results are shown in Table IV. The intent of our analysis is to identify high performing videos and therefore, Table IV contains only the videos with positive z-scores and the videos with negative z-scores are ignored.

TABLE IV: TED TALK VIDEOS WITH POSITIVE Z-SCORES

| Talk | z-score for<br>cumulative | z-score for<br>peak | z-score for<br>cumulative | z-score<br>for peak |
|------|---------------------------|---------------------|---------------------------|---------------------|
| ID   | comments                  | comments            | views                     | views               |
| 1551 | 0.26                      | 0.21                | -                         | -                   |
| 1552 | 0.03                      | -                   | -                         | -                   |
| 1554 | 2.43                      | 2.21                | -                         | -                   |
| 1555 | 1.63                      | 0.36                | 2.33                      | -                   |
| 1558 | 1.13                      | 1.39                | 0.04                      | -                   |
| 1561 | 0.09                      | -                   | -                         | -                   |
| 1564 | 1.72                      | -                   | 0.19                      | -                   |
| 1568 | 1.23                      | 1.29                | 1.11                      | 0.35                |
| 1570 | 2.13                      | 2.51                | 0.78                      | -                   |
| 1571 | 0.56                      | -                   | 1.12                      | -                   |
| 1573 | 1.44                      | 2.82                | 3.69                      | 1.98                |
| 1574 | 0.16                      | 0.69                | -                         | -                   |
| 1576 | 0.83                      | 0.90                | 1.06                      | 0.21                |
| 1580 | -                         | 0.11                | -                         | -                   |
| 1581 | -                         | -                   | 0.15                      | -                   |
| 1583 | -                         | -                   | 1.38                      | 5.17                |

In Table V, we define a benchmark at varying confidence levels (50 percent, 75 percent, 90 percent, 95 percent, and 99 percent) and use the z-scores (results in Table IV) to identify the videos that lie outside the upper boundaries. The column labelled 'Upper confidence limit (UCL)' gives the values of upper confidence limit for both the measures comments and views and also defines the z-scores that lie within the limits. However, our research interest is to identify the high performing videos, therefore, we look at z-scores falling beyond the defined upper confidence limits to identify the outliers. To identify an outlier, for example, at a 50 percent confidence level, the z-score for the videos should lie beyond +0.67 (see: Table IV). The analysis yields the same result when calculated with the upper confidence limit values. We expect the benchmarks of +0.67, +1.15, +1.64, +1.96, and +2.58 to cut off the top 25%, 12.5%, 5%, 2.5%, and 0.5% of scores respectively [21]. For our data set, we find 24% and 21% of videos had z-scores greater than +0.67 for comments and views respectively. Similarly, 21%, 12%, 9% and 3% of the videos had z-scores greater than +1.15, +1.64, +1.96, and +2.58 respectively for the measure comments. These values are higher than what would be expected in a normal distribution but is expected for our data set because of its non-normality. The similar high values are found for the measure views as well with 9% of the videos having z-scores greater than +1.15, +1.64, +1.96, and 6% of the videos having z-scores greater +2.58. This suggests that there may be too many outliers even at the higher confidence levels. In context with our research question, we decide on 75 percent confidence level to identify the high performing TED talk videos. There are seven and three TED talk videos identified as outliers for comments and views respectively (see: Table VI).

| TABLE V: IDENTIFYING HIGH PERFORMING TED TALK VIDEOS AT     |   |
|---|---|
| DIFFERENT CONFIDENCE LEVELS USING Z-SCORES FOR UPPER CUT-OF | Ŧ |

| Confidence<br>level | Upper<br>confidence<br>limit (UCL)  | Outlier TED<br>talk videos for<br>comments<br>(z-scores greater<br>than UCL) | Outlier TED<br>talk videos for<br>views<br>(z-scores<br>greater than<br>UCL) |
|---------------------|---|--|--|
| 50<br>percent       | $\bar{x}$ + 0.67 $\sigma_{\bar{x}}$<br>[Comments:<br>341.21 <sup>a</sup> ;<br>76.31 <sup>b</sup><br>Views:<br>1805219.04 <sup>c</sup> ;<br>270886.02 <sup>d</sup> ]                   | Talk IDs 1554**,<br>1555, 1558**,<br>1564**, 1568,<br>1570, 1573, 1576       | Talk IDs 1555,<br>1568, 1570,<br>1571*, 1573,<br>1576, 1583*                 |
| 75 percent          | $\bar{x}$ + 1.15 $\sigma_{\bar{x}}$<br>[Comments:<br>416.89 <sup>a</sup> ;<br>92.13 <sup>b</sup><br>Views:<br>2196429.11 <sup>c</sup> ;<br>382350.09 <sup>d</sup> ]                   | Talk IDs 1554**,<br>1555, 1558**,<br>1564**, 1568**,<br>1570**, 1573         | Talk IDs 1555,<br>1573, 1583*  |
| 90 percent          | $ \begin{split} \bar{x} + 1.64 \ \sigma_{\bar{x}} \\ \hline [Comments: \\ 494.16^a; 108.27^b \\ Views: \\ 2595789.39^c; \\ 496136.32^d] \end{split} $                                 | Talk IDs 1554**,<br>1564**, 1570**,<br>1573                                  | Talk IDs<br>1555*, 1573,<br>1583*  |
| 95 percent          | $\bar{x}$ + 1.96 $\sigma_{\bar{x}}$<br>[Comments: 544.62 <sup>a</sup> ; 118.82 <sup>b</sup><br>Views: 2856596.12 <sup>c</sup> ; 570445.69 <sup>d</sup> ]                              | Talk IDs 1554**,<br>1570**, 1573   | Talk IDs<br>1555*, 1573,<br>1583*  |
| 99 percent          | $\bar{\mathbf{x}}$ + 2.58 $\sigma_{\bar{\mathbf{x}}}$<br>[Comments:<br>642.38 <sup>a</sup> ;<br>139.24 <sup>b</sup><br>Views:<br>3361909.13 <sup>c</sup> ;<br>71.4400.13 <sup>d</sup> | Talk ID 1573   | Talk IDs 1573,<br>1583*  |

<sup>\*</sup>Outliers unique to TED video views.

\*\*Outliers unique to TED video comments.

a, b, c, d represents upper confidence limit value for cumulative

comments, peak comments, cumulative views and peak views respectively.

**Identifying outliers using box plot:** The outliers for high performing videos can also be identified using box plots. Identifying outliers using box plot requires testing each variable separately for outliers. The outliers are selected for

data points lying above the 3<sup>rd</sup> quartile (75 percentile) and the results are shown in Table VI. The 3<sup>rd</sup> quartile values are 334 cumulative comments, 62.25 for peak comments, 1389315 for cumulative views and 83207 for peak views. There are seven and eleven TED talk videos identified as outliers for comments and views respectively.

The comparison of the two methods, using z-scores and box plot, for identifying outliers is shown in Table VI. The result also shows the strength of the correlation relationship for the different variables. The box plot method for identifying outliers shows a wider range with six unique TED talk videos (Talk IDs 1550, 1571, 1574, 1576, 1581, and 1582) identified as outliers that are not captured using the z-score method. However, in contrast to the box plot method which requires separate testing of each variable to identify outliers, the z-score method standardizes the values across variables (see: Table IV) and uses a single z-score at a selected confidence level to identify outliers. This means that the z-score method can be used to evaluate different types of measurement outputs, like comments and views for a video, with different variables, on a standardized scale of z-score. The eight videos identified as outliers using z-score method are flagged for further investigation.

|  | 11112 02110   | L SCORL A                  | DOA .                                  | LOI MILII  | 1005                             |                       |
|--|---|----------------------------|--|--|----------------------------------|-----------------------|
|  |   | Strengt                    | h of cor                               |  | Outlier                          |                       |
| Method   | Outliers for<br>comments<br>and views                   | Medium<br>effect<br>(±0.3) | Large<br>effect<br>(±0.5)              | Very<br>strong<br>effect<br>(±0.75)                      | Outlier<br>count for<br>comments | count<br>for<br>views |
| Using<br>z-scores  | Cumulative  | -                          | -                                      | Talk IDs<br>1555,<br>1564**                              | 2                                | 1                     |
| for values<br>greater than   | Peak  |                            | -                                      | Talk ID<br>1558**  | 1                                | -                     |
| 75 percent<br>confidence<br>level  | Outliers<br>common to<br>peak and<br>cumulative         | Talk ID<br>1583*           | Talk<br>ID<br>1568*<br>*,              | Talk IDs<br>1554**,<br>1570**,<br>1573                   | 4                                | 2                     |
| Using box<br>plot<br>Outliers set  | Cumulative  | -                          | Talk<br>ID<br>1571*                    | Talk ID<br>1564  | 1                                | 2                     |
| Outhers set<br>for values<br>greater than<br>3 <sup>rd</sup> quartile-<br>75<br>percentile | Peak  | Talk ID<br>1582*           | -                                      | Talk ID<br>1574**,<br>1550*                              | 1                                | 2                     |
|  | Outliers<br>common to<br>both peak<br>and<br>cumulative | Talk ID<br>1583*           | Talk<br>IDs<br>1568,<br>1576,<br>1581* | Talk IDs<br>1554**,<br>1555,<br>1558**,<br>1570,<br>1573 | 7                                | 7                     |

TABLE VI: IDENTIFYING OUTLIER TED TALK VIDEOS FOR COMMENTS AND VIEWS USING Z-SCORE AND BOX PLOT METHODS

\*Outliers unique to TED video views.

\*\*Outliers unique to TED video comments.

# C. Exponential Growth Analysis Using Curve Fitting

The flagged eight high performing TED talk videos are further tested to check if they follow viral phenomenon. From the current literatures in biology and marketing, the established indicator of viral phenomenon is the exponential growth. Therefore, we check if the high performing TED talk videos follows viral phenomenon based on whether its growth trajectory follows an exponential growth pattern. We test this using curve fitting method, where a video is checked for a closer fit to exponential growth in comparison to other type of growth curves like linear, S-curve and logarithmic. The curve fitting using nonlinear least squares regression method quantifies the fit to the data points based on an evaluation of higher  $R^2$  (coefficient of determination) and lower SS residuals (sum of squares of the residuals) and MSE (mean squared value) value [22]-[24]. The decision on the set of data points to be considered for the curve fit analysis is restricted to the maximum growth phase of the curve during the entire time period, because a lack of significance in the fit will be observed if all the data is used [24]. The result of the exponential growth analysis using curve fitting method is shown in Table VII.

# D. Categorization of Viral Videos

TABLE VII: SUMMARY OF VIRAL VIDEOS BASED ON THE MEASURES COMMENTS AND VIEWS

|            | Darform         | Measure                              |         |                                  |                        |                         |
|------------|-----------------|--------------------------------------|---------|----------------------------------|------------------------|-------------------------|
| Talk<br>ID | ance<br>measure | High performing<br>videos (Outlier)‡ |         | Follows<br>exponential<br>growth | Identified<br>as viral | Strength of correlation |
| 1583*      | Comments        | No                                   | No      | -                                | -                      | Medium                  |
|            | Views           | Outlier                              | Outlier | Yes                              | Yes                    | effect $(\pm 0.3)$      |
| 1573       | Comments        | Outlier                              | Outlier | No                               | No                     | Very strong             |
|            | Views           | Outlier                              | Outlier | Yes                              | Yes                    | (±0.75)                 |
| 1570**     | Comments        | Outlier                              | Outlier | No                               | No                     | Very strong             |
|            | Views           | No                                   | No      | No                               | No                     | (±0.75)                 |
| 1568**     | Comments        | Outlier                              | Outlier | No                               | No                     | Large                   |
|            | Views           | No                                   | No      | Yes                              | No                     | effect (±0.5)           |
| 1564**     | Comments        | Outlier                              | No      | Yes                              | Yes                    | Very strong             |
|            | Views           | No                                   | No      | Yes                              | No                     | effect<br>(±0.75)       |
| 1558**     | Comments        | No                                   | Outlier | No                               | No                     | Very strong             |
|            | Views           | No                                   | No      | -                                | -                      | (±0.75)                 |
| 1555       | Comments        | Outlier                              | No      | No                               | No                     | Very strong             |
|            | Views           | Outlier                              | No      | Yes                              | Yes                    | (±0.75)                 |
| 1554**     | Comments        | Outlier                              | Outlier | No                               | No                     | Very strong             |
|            | Views           | No                                   | No      | -                                | -                      | (±0.75)                 |

*†To be categorized as viral, a video should follow exponential growth and also be a high performing video for the variables (cumulative, peak).* 

*‡ Identified outliers of high performing videos using the method of z-scores. \*Outliers unique to TED video views.* 

\*\*Outliers unique to TED video comments.

David Steindl-Rast: Want to be happy? Be grateful (Talk ID: 1555)



Fig. 1. Representation of daily views and comments (with normalized values) for TED talk video with Talk ID 1555.

Fig. 1 shows the representation of daily views and comments for a TED talk video with Talk ID 1555. Since video views are of much higher magnitude than comments, the values are normalized to represent the comments and views on the same scale. For the measurement of viral phenomenon, a TED talk video is considered a viral video when it follows an exponential growth and is also a high performing video for the variables cumulative and peak with respect to the performance measure. From Table VII, we observe that only one high performing outlier video, with Talk ID 1564, amongst the other high performing videos for the measure of video comments, follows exponential growth. Therefore, the talk is considered as following viral phenomenon for the measure comments. For the measure of video views, three videos with Talk IDs 1555, 1573 and 1583, identified as high performing videos, are categorized as following viral phenomenon because of its exponential growth. The result also shows that for the measure views, Talk IDs 1564 and 1568 also follows exponential growth. However, these talks do not emerge as outliers for the measure video views. Therefore, despite the videos following exponential growth are not considered as viral videos because they are not high performing outlier videos.

# V. DISCUSSION AND CONCLUSION

#### A. Discussion and Conclusion

In this paper, we perform a measurement study and present a novel methodology to compare different performance measures and test its validity for measuring a viral phenomenon. The methodology is empirically tested using the data of TED talk videos with two different video measures of comments and views.

We summarize our findings as follows:

(A) There is a strong correlation between the two different video output measures, namely comments and views. Our findings compliments the work of [17] in showing that there is a relation between the two video measures. However, the results also show that video views propagate much more widely and has a longer sustenance time when compared to video comments.

(B) The measure of video views better represents viral videos when compared with the measure of video comments. A video which exhibits viral phenomenon based on one measure (say views, for Talk IDs 1555 and 1573) may not follow viral phenomenon for another performance measure (comments) despite having a very strong correlation effect between the measures. When we look at the common outliers for both the measures (Talk IDs1555 and 1573), the viral phenomenon is observed only with the measure video views. This indicates that video views has a higher strength in driving viral videos when compared to the measure comments, and therefore, is a valid performance measure or indicator for measuring viral videos. The measure of video views is the valid indicator for the measurement of viral phenomenon of videos, contrary to the existing studies suggesting video comments as a viral measurement indicator.

The reason why the TED video with Talk ID 1564 emerged as following viral phenomenon with respect to comments (high performing outlier video and also followed exponential growth) but failed to be categorized as viral with respect to views (followed exponential growth but not a high performing outlier video) can be probed further in future research by analyzing the talk in the context of the comments i.e., insights gained through content analysis of comments.

#### B. Managerial Implications

Our study offers practitioners an empirical method to evaluate the different performance measures for a video and tests for the validity of the measures in driving viral phenomenon. The results have significance to business strategists and marketing managers.

The two different performance measures for the videos namely views and comments, considered in our analysis, display disparate characteristics. The video views have a longer sustenance time and are of much higher magnitude than comments. Therefore, our method can be used to investigate not just two different measures of a singular product category but also of two dissimilar product categories, say a product and a video with its different output measures like product sales and video views, and test for the viral phenomenon on a comparable basis. Our evaluation method of validating different performance measures and testing for its validity as as a better indicator in driving viral phenomenon is generalizable across product categories and market types.

The insights from identifying a performance measure as a valid viral indicator can help a business strategist in decision making on which performance measure needs business focus in driving virality. A marketing manager can channel the resources and customers towards the performance measure that drives virality.

#### REFERENCES

- [1] S. Jurvetson, "What is viral marketing?" *Draper Fisher Jurvetson (DFJ) Network News*, 2000.
- [2] A. M. Kaplan and M. Haenlein, "Two hearts in three-quarter time: How to waltz the social media/viral marketing dance," *Business Horizons*, vol. 54, no. 3, pp. 253-263, 2011.
- [3] P. Grifoni, A. D. Andrea, and F. Ferri, "An integrated framework for on-line viral marketing campaign planning," *International Business Research*, vol. 6, no. 1, pp. 22-30, 2013.
- [4] J. Hemsley, "Virality: Developing a rigorous and useful definition of an information diffusion process," *Communication and Social Networks*, University of Washington, 2011.
- [5] K. Nahon, J. Hemsley, S. Walker, and M. Hussain, "Fifteen minutes of fame: The power of blogs in the lifecycle of viral political information," *Policy & Internet*, vol. 3, no. 1, article 2, 2011.
- [6] S. Jurvetson and T. Draper, "Viral marketing: Viral marketing phenomenon explained," *Draper Fisher Jurvetson (DFJ) Network News*, 1997.
- [7] A. L. Montgomery, "Applying quantitative marketing techniques to the Internet," *Interfaces*, vol. 31, no. 2, pp. 90-108, 2001.
- [8] K. Lynch. (March 2013). World book day: Our top ten readingrelated world records. [Online]. Available: http://www.guinnessworldrecords.com/news/2013/3/world-book-day-o ur-top-ten-reading-related-world-records-47374
- [9] J. Berger, Contagious: Why things catch on, Simon & Schuster, 2013.
- [10] C. Briggs, "BlendTec will it blend? Viral video case study," *SocialLens. Com*, pp. 1-8, January 2009.
- [11] K. Wallsten, "Yes we can: How online viewership, blog discussion, campaign statements, and mainstream media coverage produced a viral

video phenomenon," *Journal of Information Technology and Polictics*, vol. 7, no. 2-3, pp. 163-181, 2010.

- [12] D. Cruz and C. Fill, "Evaluating viral marketing: Isolating the key criteria," *Marketing Intelligence Planning*, vol. 26, no. 7, pp. 743-758, 2008.
- [13] K. Kalyanam, S. McIntyre, and J. T. Masonis, "Behind the scenes of a viral marketing campaign: How plaxo crossed the tipping point and avoided the fate of the ebola virus," Rapport Technique, Santa Clara University, Santa Clara, California, vol. 40, 2007.
- [14] J. Rayport. (December 1996). The virus of marketing. Fast Company. [Online]. Available: http://www.fastcompany.com/27701/virus-marketing
- [15] Y. Liu-Thompkins, "Seeding viral content: The role of message and network factors," *Journal of advertising research*, pp. 59-72, December 2012.
- [16] R. Ferguson, "Word of mouth and viral marketing: Taking the temperature of the hottest trends in marketing," *Journal of Consumer Marketing*, vol. 25, no. 3, pp. 179–182, 2008.
- [17] R. Zhou, S. Khemmarat, and L. Gao, "The impact of YouTube recommendation system on video views," in *Proc. the 10th ACM SIGCOMM Conference on Internet Measurement*, ACM, pp. 404-410, 2010.
- [18] M. Cha, A. Mislove, and K. P. Gummadi, "A measurement-driven analysis of information propagation in the Flickr social network," in *Proc. the 18th International Conference on World Wide Web*, ACM, pp. 721-730, 2009.
- [19] T. Broxton, Y. Interian, J. Vaver, and M. Wattenhofer, "Catching a viral video," *Journal of intelligent information systems*, vol. 40, no. 2, pp. 241-259, 2013.
- [20] J. Cohen, "A power primer," *Psychological Bulletin*, vol. 112, no. 1, p. 155, 1992.
- [21] A. Field, Discovering Statistics Using SPSS, Beverly Hills: Sage publications, 2009.
- [22] A. M. Brown, "A step-by-step guide to non-linear regression analysis of experiment data using a microsoft excel spreadsheet," *Computer Methods and Programs in Biomedicine*, vol. 65, no. 3, pp. 191-200, 2001.
- [23] H. J. Motulsky and L. A. Ransnas, "Fitting curves to data using nonlinear regression: A practical and nonmathematical review," *The FASEB Journal*, vol. 1, no. 5, pp. 365-374, 2012.
- [24] H. Motulsky and A. Christopoulos, *Fitting Models to Biological Data Using Linear and Nonlinear regression A Practical Guide to curve Fitting*, San Diego, CA: GraphPad Software Inc., Oxford University Press, 2003.



Archana Anand Boppolige is a research scholar at the Department of Management Studies, Indian Institute of Science, Bangalore, India. She has received a master"s degree in Master of Business Administration from Indian Institute of Technology, Kharagpur, India. She has also worked as a Project Manager and an Application Consultant at IBM. Her area of research is viral marketing.



**Anjula Gurtoo** is Associate Professor at the Department of Management Studies, Indian Institute of Science, Bangalore, India. She has published more than 35 journal papers in the area of public policy. She has received the DST Pt. Jawaharlal Nehru National Award (2010-12) and the Social Science Research Award (2009) for overall contribution to research. Her two recent books are Handbook of Entrepreneurship in Developing Economies, Springer, United Kingdom,

and Developing Country Perspectives on Public Service Delivery, Springer, New Delhi. She has managed research projects from the Ford Foundation USA, World Resources Institute, Washington D.C., USA, British Council of India, Wage Indicator Foundation, The Netherlands, and United States Agency for International Development (USAID), Washington D.C, USA.