Team 1153

IF A NEW ZEALAND STUDENT UPLOADS A VIDEO CLIP THAT GOES VIRAL, HOW LONG WILL IT TAKE BEFORE 1% OF THE WORLD'S POPULATION HAS SEEN IT?

Summary:

Viral videos are characterised by gaining large numbers of views in short periods of time. While most viral videos only gain between 5 and 40 million views, thus making it unlikely that such a video will ever be seen by 1% of the world's population, a few viral videos do reach significantly higher numbers of views, upwards of 100 million views. Our definition of viral, taken from a well-known YouTube celebrity, says that a video must gain 5 million views in 3-7 days in order to be considered viral. This led to the result that for 1% of the world's population to see a video clip uploaded by a New Zealand student that has gone viral, a period of time ranging from 8.63 days to 18.81 days was required after it was uploaded.

After defining what we meant by viral videos and views we went on to model the world population in order to determine exactly what constitutes 1% of the world population at the time after being uploaded when the video would have sufficient views for this. We then modelled the way viral YouTube videos gain views using the Gompertz model, which is essentially an improvement of the logistic model for growth that deals with the fact that views gained by viral videos does not tend to be symmetrical in the convex and concave portions of the graph. We figured out the maximum amount of people that would be likely to watch a high-quality viral video to be 100,000,000 and used this as our maximum value in the Gompertz model.

Introduction:

We are interpreting the question to mean how long from the video's upload will it take before 1% of the world's population has seen the video. We assume that the video is uploaded at 1pm on 31st of July 2015, one day before our baseline for the world population model in order to account for our assumption that in the first 24 hours after uploading the video we will only gain 200 views and not consider the video viral, however the results would be very similar for videos uploaded a few weeks earlier or later than this. The definition of a 'viral video' according to popular YouTube celebrity Kevin Nalty is a video that gets more than 5 million views in a three to seven day period.[1] We assume that the video has already gone viral, and therefore gains the required 5 million views. This means we do not have to focus on how the video will get viral or its ability to do so. We are also assuming that the video is funny or interesting enough to gain at least 1% of the world's population's views. In reality this is unlikely as very few viral videos actually gain these many views.

We are also ignoring music videos as these videos are not viral in the full sense of the word, and are instead just popular. Also many views on music videos are from individuals listening to the song multiple times.

The video being uploaded by a New Zealand student has little effect on the speed at which views will be gained. This is because we live in an extremely well connected world and international views are easily gained through the sharing that occurs on websites such as Reddit, Twitter and Facebook. These social media devices are essential in promoting videos to viral status, and thus a video that has gone viral will almost definitely be shared vigorously on at least two of these websites. Therefore we can assume that the origin of the video being New Zealand will have little effect on its speed of gaining views.

[1] http://www.adweek.com/socialtimes/what-makes-a-video-viral/62414

What constitutes a view?

Viral videos that achieve the number of views required to satisfy 1% of the world's population are found almost exclusively on YouTube. The next largest video-sharing site, Vimeo, has only 68.2 million on its most viewed video "The Mountain" by TSO Photography.[2] Therefore we can assume that the number of people who have seen the video is given by the view-count on YouTube. YouTube's view-count does allow a single user to view a video multiple times[3], however in the case of most viral videos, a viewer will not watch it multiple times alone, as these videos lose their appeal quickly. However it is likely that they may show their friends, in which case the extra view that YouTube counts will be accounted for by the extra person watching the video. Depending on the number of friends the video is shown to at once, or depending on the number of times an individual views the video can cause over or under-estimates of the time required for 1% of the world's population to view the video, but based on the low re-watch appeal of most viral videos we will assume that 1 YouTube view is equal to one unique individual watching the video.

World Population:

We will assume that the video that is going viral has been uploaded today. The world population at 1pm on the 1st of August 2015 is 7,356,817,000 people and the current population growth is 1.13% yearly which equates to 0.00309589041% a day.[4] This rate of population growth is expected to decrease, however we only need to look at a short period of time past the current date as a video will be gaining views fairly quickly if it is viral. We can therefore use this rate of increase as an accurate estimate of population growth in a basic exponential model of growth. We have chosen not to use the logistic model of growth to model world population growth as we have accurate information for the current population growth as a percentage, and since we are looking at a very short time-frame, this rate of growth will hold. If we were to use the logistic model, we would have to estimate the maximum carrying capacity of the Earth, which would lead to the potential for much larger inaccuracies.

The general equation for the exponential model we are using is given by:

$$P(t) = P_0(1 + r)^t \qquad (1)$$

Where P is the population t days after some base time, P_0 is the population at said base time and r is the rate by which the population increases yearly. Therefore we model the world population by the following equation, with $P_0 = 7{,}356{,}817{,}000$ and r = 0.0000309589041, which will give accurate estimates of the world population over the next few months.

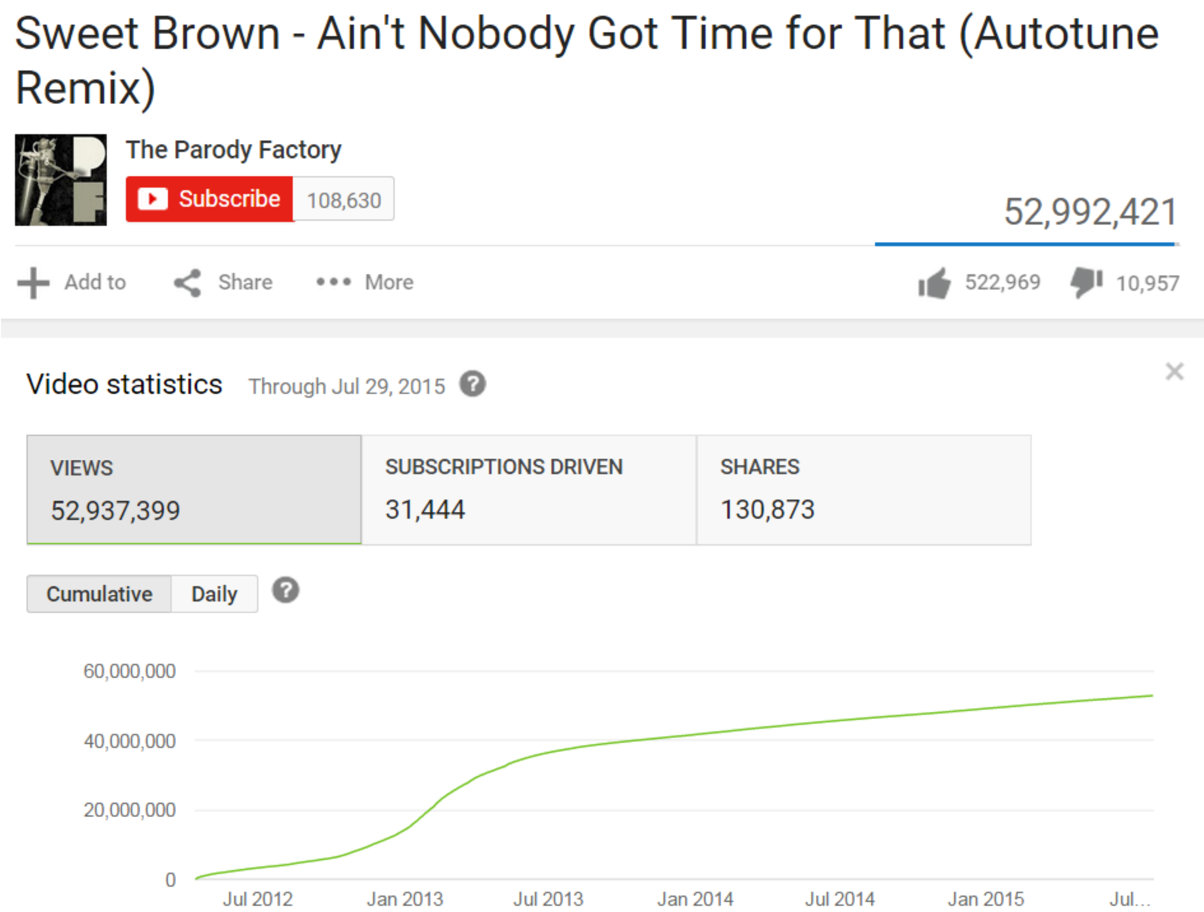
$$P(t) = 7{,}356{,}817{,}000(1.0000309589041)^t \qquad (2)$$

[2] http://www.vimeo.com/channels/top
[3] http://www.bluefountainmedia.com/blog/how-does-youtube-count-views/
[4] http://www.worldometers.info/world-population/

P is the world population in number of people, and t is the time elapsed in days from the base time of 1pm on the 1st of August 2015. We will use this result later on once we have modelled the growth of views on a viral video in order to pinpoint what we mean by 1% of the population.

Modelling YouTube View Growth:

To model the growth of YouTube views we considered that a video is likely to have a maximum amount of views it will gain, even after a very long time period. This was supported by the statistics we saw on YouTube videos that showed us that most viral videos approached an upper limit, which depended on the type of video in question. For a high-quality viral video it seemed that 100,000,000 views was a reasonable estimate. The statistics also showed us view growth was S-shaped, and therefore a sigmoid model curve would be the most appropriate model. The growth was not symmetrical however, and this discouraged us from using the logistic model for growth. The picture below shows a popular viral video and the growth of views as an asymmetrical sigmoid curve.



Source: https://www.youtube.com/watch?v=bFEoMO0pc7k

Due to the asymmetry of the above curve, we decided to use a Gompertz model as this deals with the symmetry problem associated with the logistic model for growth.[5] The Gompertz model is as follows:

[5] http://arxiv.org/pdf/1404.2570.pdf

$$S(t) = Me^{-\ln\left(\frac{M}{S(0)}\right)e^{-\lambda t}} \tag{3}$$

In this model S(t) is the number of views after t days, M is the maximum number of views the video is expected to receive, and λ is a constant related to how quickly the video gains views. Looking at other highly viral videos on YouTube we have noticed that 100 million views is approximately where the number of views tends to level out at.[6] We also decided that this model is only particularly accurate once a video has begun to go viral. This is very unlikely to happen in the first 24 hours after uploading the video. The only views occurring at this point in time will be from the uploader's direct network, such as family and Facebook friends. This is likely to bring in approximately 200 views in the first day based on the first-day statistics of most viral videos. The t in S(t) therefore represents time in days such that t = 0 is the end of the first day after the video was uploaded. Therefore we used the Gompertz model, with M = 100,000,000 and S(0) = 200. After the first day we can say that the video goes viral and will gain its required 5 million views in the following 3 to 7 days. So in order to find the value of λ for a viral video that gains 5 million views in the next 3 to 7 days we can set S(t) = 5,000,000 and t = 3 and solve for λ and then repeat the process with t =7. This will give us the range of λ values for a video that satisfies the view growth conditions of a viral video as outlined by our definition of a viral video. The equations we have to solve are:

$$S(3) = 100{,}000{,}000e^{-\ln\left(\frac{100{,}000{,}000}{200}\right)e^{-\lambda t}} \tag{4}$$

$$S(7) = 100{,}000{,}000e^{-\ln\left(\frac{100{,}000{,}000}{200}\right)e^{-\lambda t}} \tag{5}$$

Solving (4) we get λ ≈ 0.4924 and solving (5) we get λ ≈ 0.2110. Therefore we can use these two values for lambda to find boundary conditions for how long it will take a viral video to be seen by 1% of the world's population. Therefore we have two equations depending on the value of lambda

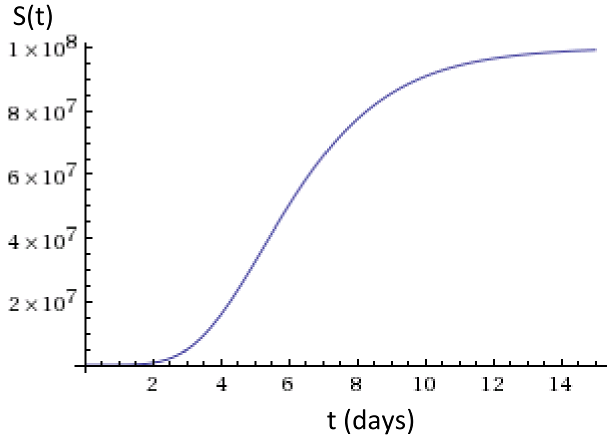
$$S_1(t) = 100{,}000{,}000e^{-\ln\left(\frac{100{,}000{,}000}{200}\right)e^{-0.4924t}} \tag{6}$$

$$S_2(t) = 100{,}000{,}000e^{-\ln\left(\frac{100{,}000{,}000}{200}\right)e^{-0.2110t}} \tag{7}$$

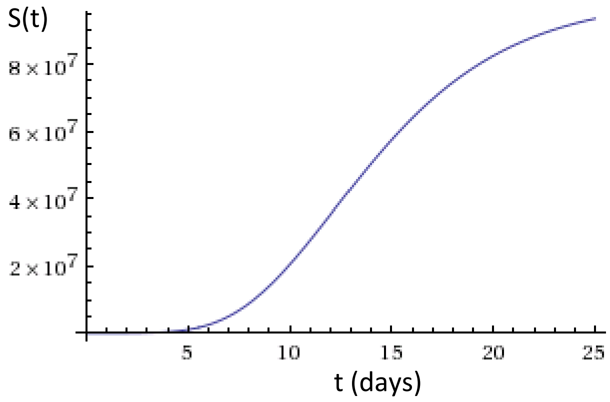
The minimum time required to reach 1% of the world's population will be achieved by equation (6) and the maximum time required will be achieved by equation (7). Graphs of both are drawn below.

[6] https://www.youtube.com/watch?v=3XviR7esUvo, https://www.youtube.com/watch?v=YoB8t0B4jx4, https://www.youtube.com/watch?v=Y4MnpzG5Sqc

Graph of $S_1(t) = 100,000,000e^{-\ln\left(\frac{100,000,000}{200}\right)e^{-0.4924t}}$



Graph of $S_2(t) = 100,000,000e^{-\ln\left(\frac{100,000,000}{200}\right)e^{-0.2110t}}$



The functions, $S_1(t)$ and $S_2(t)$ tells us the views gained by a viral video in t days. $S_1(t)$ gives us the minimum number of days required for a video that is considered viral to gain a certain number of views, while $S_2(t)$ gives us the maximum number of days required.

We can now look at what value we want $S_1(t)$ and $S_2(t)$ to take. We want to know how many days it will take for 1% of the world's population to have seen it. We can look back at equation (2) as this gives us the world's population t days after the same base time as equations (6) and (7). It's important to remember that our usage of the Gompertz model starts when the video goes viral, which we are assuming to be 24 hours after it is uploaded. Therefore the baseline time for the population growth model is the same baseline time as the Gompertz models. We want 1% of the world's population to have seen the video, and this is equivalent to 0.01P(t), therefore we can equate 0.01P(t) with $S_1(t)$ and $S_2(t)$ and solve to get minimum and maximum values for the amount of time it will take till 1% of the world population see a viral video.

For the minimum time required:

$$0.01P(t) = S_1(t) \tag{8}$$

Substituting P(t) and $S_1(t)$ into equation (8) we get:

$$0.01 * 7{,}356{,}817{,}000(1.0000309589041)^t = 100{,}000{,}000e^{-\ln\left(\frac{100{,}000{,}000}{200}\right)e^{-0.4924t}}$$

We then rearrange the above so that the right hand side equals 0 and we can use Newton's method.

$$(1.0000309589041)^t - \frac{100{,}000{,}000}{73{,}568{,}170}e^{-\ln\left(\frac{100{,}000{,}000}{200}\right)e^{-0.4924t}} = 0$$

We then create a function $f_1(t_n)$ which has the formula above.

$$f_1(t_n) = (1.0000309589041)^t - \frac{100{,}000{,}000}{73{,}568{,}170}e^{-\ln\left(\frac{100{,}000{,}000}{200}\right)e^{-0.4924t}}$$

By Newton's method, which is:

$$t_{n+1} = t_n - \frac{f_1(t_n)}{f_1'(t_n)}$$

And using an initial guess of $t_n = 10$ days, we obtain the answer:

$$t \approx 7.63\ days \tag{9}$$

This result is the minimum time required for 1% of the world's population to see the video after the initial 24 hour period where the video has not yet gone viral. Note that we are ignoring the first 200 views the video receives as it is negligible compared to 1% of the world's population.

For the maximum time required:

$$0.01P(t) = S_2(t) \tag{10}$$

Substituting P(t) and $S_2(t)$ into equation (10) we get:

$$0.01 * 7{,}356{,}817{,}000(1.0000309589041)^t = 100{,}000{,}000e^{-\ln\left(\frac{100{,}000{,}000}{200}\right)e^{-0.2110t}}$$

We then rearrange the above so that the right hand side equals 0 and we can use Newton's method.

$$(1.0000309589041)^t - \frac{100{,}000{,}000}{73{,}568{,}170}e^{-\ln\left(\frac{100{,}000{,}000}{200}\right)e^{-0.2110t}} = 0$$

We then create a function $f_2(t_n)$ which has the formula above.

$$f_2(t_n) = (1.0000309589041)^t - \frac{100{,}000{,}000}{73{,}568{,}170}e^{-\ln\left(\frac{100{,}000{,}000}{200}\right)e^{-0.2110t}}$$

By Newton's method, which is:

$$t_{n+1} = t_n - \frac{f_2(t_n)}{f_2'(t_n)}$$

And using an initial guess of $t_n = 20$ days, we obtain the answer:

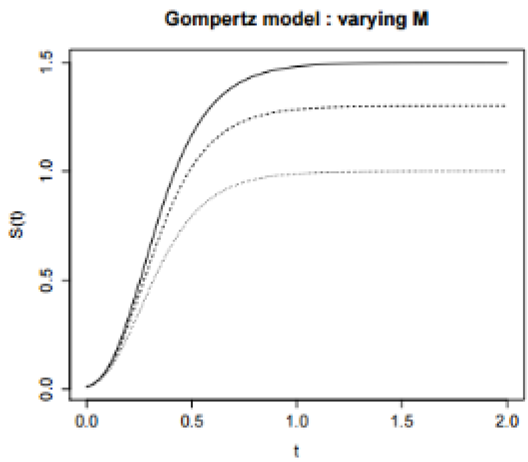
$$t \approx 17.81\ days \qquad (9)$$

This result is the maximum time required for 1% of the world's population to see the video after the initial 24 hour period where the video has not yet gone viral. We are again ignoring the first 200 views the video receives as it is negligible compared to 1% of the world's population.

As t is the number of days required after the video has gone viral for 1% of the world to see it, the actual number of days required will be t + 1. This gives us a range from a minimum of 8.63 days to a maximum of 18.81 days for 1% of the world's population to see the viral video uploaded by an NZ student.

Evaluations of Model and Assumptions:

Our main assumption was during the Gompertz model, when we estimated the maximum number of views that the viral video clip would get. Many videos on YouTube which have gone viral have ended up with wildly different numbers of views, ranging from 5 million to over 800 million. It depends on factors such as genre of the video (parody, original creation, comedy sketch, news item), how funny and interesting the video itself is and how much coverage the video receives. For example, a video that has been retweeted by Katy Perry, the most followed celebrity on twitter would receive far more views in total than a similar video that had not been retweeted by her. Varying the total number of views a video receives over a long period of time can have a reasonable impact on the accuracy of our predictions. The effect of varying M, the maximum views the video receives. Is shown on the graph below.



Source: http://arxiv.org/pdf/1404.2570.pdf

The difference between a video that receives 100,000,000 views in total and one that receives 150,000,000 is noticeable when comparing how quickly they are seen by 1% of the world's population. If we had more time to investigate this it is likely that we would have used different maximum numbers of views with the Gompertz model as we expect a higher number of maximum views would mean the video would be seen by 1% of the world's population faster. We did decide that 100,000,000 views was an appropriate maximum number of views to choose as there were a few viral videos that achieved approximately 100,000,000 views, for example the Kony 2012 video.[7] This was a good choice for this as it went viral fairly quickly and was likely only watched once due to its length and genre (documentary).

This leads on to our second assumption, which was that the video is only watched once by each person. This is a fairly reasonable assumption as people don't often re-watch viral videos, as most of the humour is gained from the first viewing. However it is possible that they show their friends in person, which can lead to only one view being counted for 4 or 5 people (or more). This could lead to an over-estimate in the amount of time required. Equally, one person may watch the video multiple times and this is more likely for shorter, funnier clips, so it is possible that by assuming each person watching the video once we are actually under-estimating the amount of time required. Therefore these under and over-estimates are likely to even out, and it is a fair assumption to make.

The use of the Gompertz model itself had the potential to lead to accuracies. A YouTube video is unlikely to have a set maximum number of views it will ever reach, and after the initial sigmoid shape of the curve it is more likely to continue upwards at a slow rate, similar to the end of a logarithmic curve. This would likely not affect our results too much, as a viral video with 100,000,000 maximum views would be seen by 1% of the world's population while still in the middle of the S-shape, near enough to the point of inflection that it would not be affected by the levelling off of the curve as it approaches 100,000,000. Therefore the use of the Gompertz model is a fairly accurate model for YouTube view growth.

For our value of S_0, the number of views the video had one day after its date of upload, we used 200 views. This was before applying the Gompertz model. This was our estimate of how many views we expected a video to have in its first 24 hours of being online, as it is likely that only online friends and family would see it initially. As it is viral video material, we estimated around 200 within the first 24 hours as most family and friends would share the video with their respective online friends.

We are confident in our estimate, as 200 is approximately the median number of Facebook friends.[8] We have also assumed that the uploader does not have any sort of pre-existing fan base, meaning that they would not get 301+ views in a matter of minutes. We can safely assume that around 200 people will see it initially even if not all friends watch it, due to sharing by other friends. This initial value is required as a video is extremely unlikely to be heavily shared by celebrities and on websites such as Reddit, Twitter and YouTube in its first day of upload, as it takes that much time to be noticed.

The last assumption we made was when modelling population growth. We felt like it was a fair assumption as a 1.13% yearly increase in population is expected for the next few months at least, and our data only required time values of less than 20 days.

[7] https://www.youtube.com/watch?v=Y4MnpzG5Sqc
[8] http://www.pewresearch.org/fact-tank/2014/02/03/6-new-facts-about-facebook/

Conclusion:

While the majority of viral videos are not viewed by even 1% of the world's population, there are exceptions to this. If we assume that the video created by the student has been popular enough to go viral then it will have been seen by 1% of the world's population within 8.63 and 18.81 days, based on a first 24 hour view-count of 200 and a maximum view-count of 100,000,000. The range of days required for the video to reach the 1% of the world population occurs because the video may gain the 5,000,000 required views to be considered viral in anywhere from 3 to 7 days. We used an exponential growth model called the Gompertz model as it took into account the maximum amount of views a YouTube video would receive and had a sigmoid shape similar to the logistic model of growth. It had the advantage over the traditional logistic model as it was asymmetrical, showing a more accurate depiction of YouTube view growth. It is likely that if we had chosen a higher maximum view count, less time would be required to reach 1% of the world's population and if we had chosen a smaller maximum amount of views then it would take longer, and possibly never even reach 1% of the world's population as the population might increase faster than views on the video did.