1020

"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?"

Introduction and Summary:

Internet and social networking sites are now omnipresent within our society, becoming an indelible aspect of our lives. Because of this widespread nature, the internet is able to facilitate the spreading of information, allowing files, images and videos to spread across the globe with an unprecedented speed. Recently, with the aid of YouTube and Facebook, many videos became 'viral' and reached more than millions of people. After conducting research and mathematical modelling, it was found that it would take approximately **58. 88** days for a viral video made by NZ student to reach 1% of the world (around 70 million people).

There is a large amount of complexity involved within this problem and we have simplified it into smaller stages:
1. Initial upload of video → Gains popularity and appear on YouTube NZ featured page
2. YouTube NZ featured page → International YouTube front page

Within these three different stages, we have developed different mathematical models as well as research data to predict the behavior of the internet. To strengthen our model, a computer simulation has also been written by us, providing a simulator of humans' interactions with each other as well as the time it would take for the viral video to reach different number of people through a propulsion algorithm.

Other external influences such as news reports, articles etc have also been taken into consideration for validity of our calculation.

Daily View count of a typical viral video on YouTube in first week
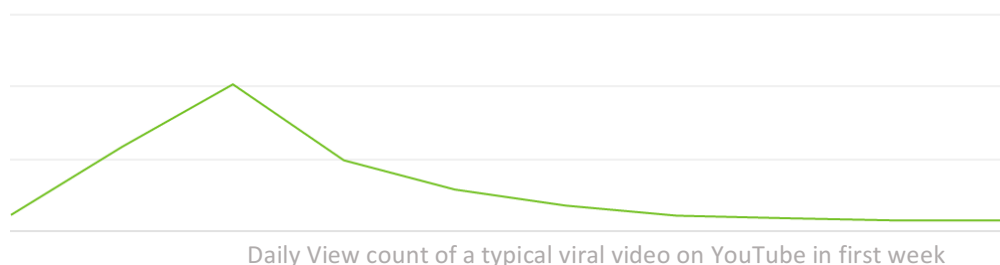
Definition of Problem
The open-ended nature of the question requires strict definitions for guidance.

"New Zealand student"
A high school or University student within the country who have little to no subscribers on YouTube but has an active Facebook account with around 338[1] participating friends. This number is the mean number of friends for users on Facebook.

"Video clip"
A 4.00-minute video clip. The time length of 4.00-minutes is calculated through finding the median of all the "popular now" videos on the New Zealand YouTube page[2].

"World population"
Defined to be 7,356,840, 000[3].

Stage 1: Initial Upload
At first, the diverse number of video sharing platforms seemed to make accurate calculations impossible. Additionally, many countries have their own dominant video sharing platform, for example China uses LeTV and PPTV rather than YouTube. However, because of the definition of the problem, it is reasonable to assume the student upload to YouTube – this is the most dominant video sharing site in NZ at 55.8%[4] of the population and there are no other platforms around the world that support English and has as much popularity. Furthermore, the video will most likely be promoted through Facebook initially.
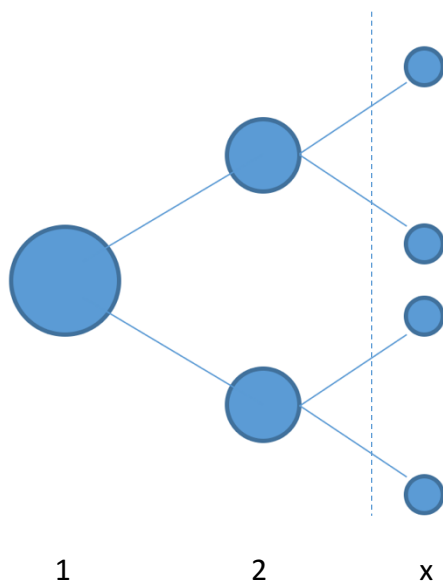
From this, we can simplify the sharing and promotion to the diagram as the diagram on the left. Each "node" or circle represents one person and each of the lines represent a connect via Facebook. Using this logic, the problem can be firstly approximated with a tree diagram.

1   2   x

---

[1] Statistics. Retrieved 01/08/2015 from http://www.pewresearch.org/fact-tank/2014/02/03/6-new-facts-about-facebook/

[2] Statistics. Retrieved 01/08/2015 from https://www.youtube.com/channel/UCYzLNwe0c5S5lnc5j3ERvFg
[3] World population. Retrieved 01/08/2015 from http://www.worldometers.info/world-population/
[4] Market Share. Retrieved 0108/2015 http://www.roimarketing.co.nz/articles/social-media-marketshare-demographics-zealand/

Initial Mathematical Modelling
This is the first mathematical model which we derived. The number of friends multiplied by the probability of a person sharing will give the number of people who can see it after one layer of branches. Putting this to the power of x, we can find out how many people watched the video and then shared after x layers of iterations.

$$Number\ of\ people\ watched\ and\ shared\ after\ x\ layers = (Friends \times \frac{Share}{Views})^x$$

However, this is rather erroneous since we are assuming that each share is to a completely different group of people which is not true since there are always mutual friends. To counter this, a coefficient $- P -$ was introduced.

$$Number\ of\ people\ watched\ and\ shared\ after\ x\ layers = (Friends \times P \times \frac{Share}{Views})^x$$

The Coefficient P
Each person's viewing affects the other. If the share is doubled up and it is "coupled"(meaning they have mutual friends), which needs to be discounted. Therefore a more sophisticated model needs to be used. First we consider a local mean field (LMF) for an arbitrary person. Each person has the average of 338 friends. It is easiest to produce a scale-free graph rooted to the arbitrary person (r). For this model, we are assuming that there are infinite amount of users, as the users on facebook/youtube is around 10^6, so this assumption is valid.

Probability of having k friends:
$$P_0 \propto k^{-\gamma}$$
And we set a constant C so that:
$$\text{Prob}[\deg(r) = k] = P_0 = Ck^{-\gamma}$$
Where:
$$C \sum_{k=1}^{\infty} k^{-\gamma} = 1$$
because all the sum of the probabilities of the branches must equal to one.
Let us denote
$$\zeta(\gamma) = \frac{1}{C}$$
Therefore substitute and rearrange to give:
$$P_0(k) = \frac{k^{-\gamma}}{\zeta(\gamma)}$$
For the a friend of the arbitrary person P_1:
$$P_1(k) = \frac{k^{-(\gamma-1)}}{\zeta(\gamma-1)}$$
as it is a power law function. This means the Riemann zeta function is:
$$\zeta(\gamma) = \sum_{k=1}^{\infty} k^{-(\gamma-1)}$$
and
$$\zeta(\gamma) = \frac{1}{\Gamma(\gamma)} \int_0^{\infty} \frac{x^{\gamma-2}}{e^x - 1} dx$$
Where $\Gamma(\gamma)$ is a gamma function:

$$\Gamma(\gamma) = \big((\gamma - 1) - 1\big)!$$

γ is a positive constant necessary to obtain to calculate P. This requires data generated. So we wrote a program in python to do so.

```python
1   import random
2   from networkx import *
3
4   #These are the input values which can be changed
5   probability_of_sharing = 0
6   number_of_friends_per_person = 0
7   number_of_people = 0
8   rewire_probability = 0.25
9   depth = 5
10
11  g = watts_strogatz_graph(number_of_people, number_of_friends_per_person, rewire_probability)
12
13  nodes = []
14
15  class Traverser:
16      def __init__(self):
17          self.nodes = []
18
19      def traverse(self, max_depth, node):
20          self._traverse(max_depth, node, 0)
21          n = self.nodes
22          self.nodes = []
23          return n
24
25      def _traverse(self, max_depth, node, depth = 0):
26          if depth >= max_depth: return
27
28          # find all adjacent nodes
29          adjacent = [x for x in g[node].keys() if x not in self.nodes]
30          self.nodes.extend(adjacent)
31          # Use random to evaluate if they share the video
32          for n in adjacent:
33              if random.random() <= probability_of_sharing:
34                  # share.
35                  self._traverse(max_depth, n, depth + 1)
36
37
38  traverser = Traverser()
39  print len(traverser.traverse(depth, 0))
```

This code also fixes the assumption that there are no inter-relationships between people, instead presents people in a more realistic way of interdependence. As shown below.
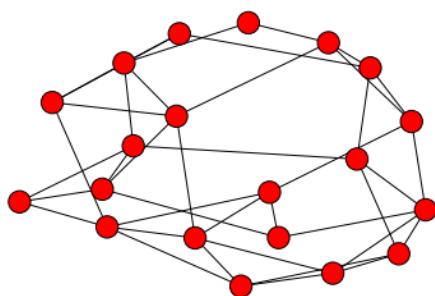
Through comparing values, γ is approximately 3.

The equation simplifies to:

$$P_0(k) = \frac{k^{-\gamma}}{\frac{1}{(\gamma-1)!}\int_0^\infty \frac{x^\gamma}{e^x-1}dx}$$

$$P_1(k) = \frac{k^{-(1-\gamma)}}{\frac{1}{(\gamma-2)!}\int_0^\infty \frac{x^{\gamma-2}}{e^x-1}dx}$$

$$P_1(k) = \frac{k^{-2}}{\frac{1}{(1)!}\int_0^\infty \frac{x}{e^x-1}dx}$$

However we only have an average of 338 friends per person:

$$P_1(k) = \frac{k^{-2}}{\frac{1}{(1)!}\int_0^{338} \frac{x^{-2}}{e^x-1}dx}$$
$$= (1.4493\times338)^{-1}$$
$$= 2.04\times10^{-3}$$

is the proportion that will send to a specific friend independently, so:
$$P_1(k)^2 = 4.16\times10^{-6}$$

is the proportion that two people will send to the same person

So to modify the previous equation of

We times FRIENDS by the amount that is NOT repeated:
$$\text{Views} = (F\times(1-4.16\times10^{-6})\times\frac{S}{V})^x$$
$$70,000,000 = (338\times(1-4.16\times10^{-6})\times\frac{220}{50000})^x$$

Solve for x
$$x = 34$$
As time from upload increases the time of iterations increase as people become less interested in the video. Let's assume that average iteration length in an infinite amount of time is 1 day so:
$$\text{Time} = 34 \text{ days}$$

The overall time for the video to reach the featured title page is therefore 34 days.

Stage2: Promotion on Featured page

The modelling after the video has arrived onto the YouTube featured site is much simpler than the initial diffusion approach. The cycle of a viral video has been extensively researched.

Social Diffusion Curve

Days Following Launch
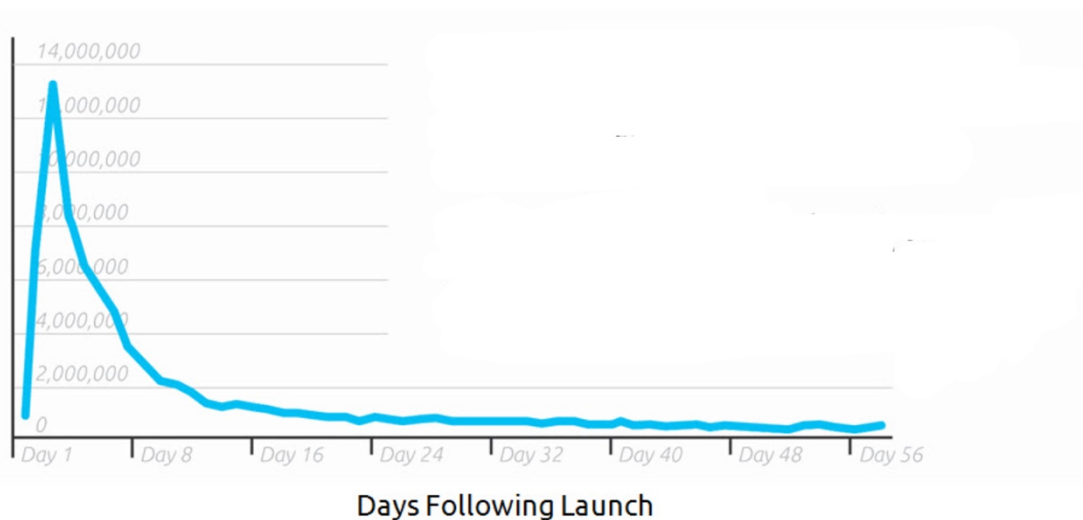
5

The graph above shows the average number of views per day for a viral video. It is reasonable to assume that this viral video also follows the paths, meaning the peak of its popularity will be reached on its fourth day after its initial appearance on the YouTube featured page. Because we are mostly interested in the spreading, we will discard the decaying part. By integration, we can find the area – cumulative number of people who has viewed. To integrate this curve, we have used graphmatica to generate two approximate graphs and separated the equations.

---

5 Graph. 01/08/2015 from http://www.businessinsider.com.au/viral-video-lifecycle-2013-4
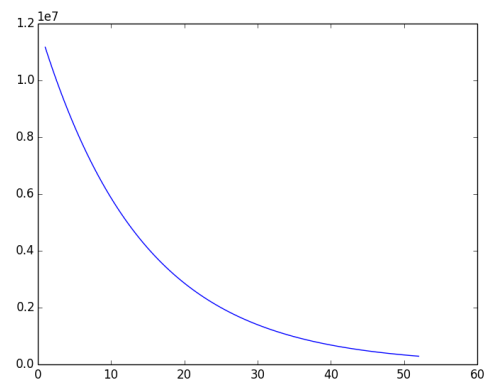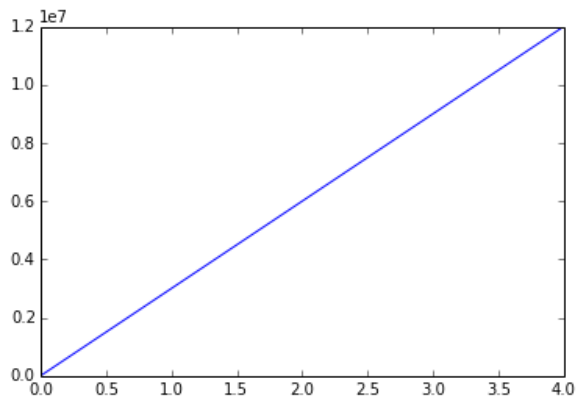
$$70 = \int_0^4 3x \, dx + \int_4^t 12e^{-0.07168x} dx$$

$$70 = 3\left[\frac{x^2}{2}\right]_0^4 - 12\left[\frac{e^{-0.071688x}}{-0.071688}\right]_4^t$$

$$70 = 24 - 12\left[\frac{e^{-0.071688x}}{-0.071688}\right]_4^t$$

$$t = 23.88$$

Overall the total time is 57.88 days.

Assumption

1. All of the initial promotion is done through Facebook
2. There are no views from YouTube initially
3. Approximation using graphmatica

Evaluation and Conclusion

Choices are very difficult to determine. Our models assumed that all friends will be sent the video (not blocked friends) and at the initial stage there are no other views from searching YouTube. This is reasonable as in videos uploaded by a little known user has a negligible amount of views relative to the "viral" status. Also the approach to uploading and sharing the video can be different.

Not all viral videos follow the "viral video view trend". There are often peaks and troughs during the social diffusion stage that may influence the views. However it is impossible to predict when these will happen as they are attributed to news articles, national news spotlights and Facebook page featuring. To further improve this, we could take into account the average viral videos and the effect of a news article on it. However, this should only effect the result by a couple of days.

Our mathematical models are only models, but they can vary from the real situation. If we are given more time we would firstly investigate other aspects of the video such as the genera etc.

Bibliography

- Equation. Retrieved 01/08/2015 from
  http://netecon.seas.harvard.edu/NetEcon09/Papers/Zhao_09.pdf

- Statistics. Retrieved 01/08/2015 from
  http://www.pewresearch.org/fact-tank/2014/02/03/6-new-facts-about-facebook/

- Statistics. Retrieved 01/08/2015 from
  https://www.youtube.com/channel/UCYzLNwe0c5S5lnc5j3ERvFg

- World population. Retrieved 01/08/2015 from
  http://www.worldometers.info/world-population/

- Market Share. Retrieved 0108/2015
  http://www.roimarketing.co.nz/articles/social-media-marketshare-demographics-zealand/

- Graph. 01/08/2015 from http://www.businessinsider.com.au/viral-video-lifecycle-2013-4