Viral spread of content on social media

Aditya Tripathi
Department of Computer Science
Indian Institute of Technology
aditya.tripathi@iitgn.ac.in

Abstract

Viral video are the videos that become popular through sharing via the internet. The sharing of videos on social sites has given rise to this phenomenon. In this paper we seek to understand the spread of viral content on social media via the methods of diffusion of information, and try to answer question like - What kind of social network graph has what kind of effect on the sharing and virality of a video? Which model is the real world social graph closest to? What factors (social and non-social) contribute to a virality of content/video? what pushes people to share watch the content? What interactions between content and user compels the users to watch the it? etc.

The model is based on real-world phenomenons. found in the real world social media networks.We try to find the viewership and sharing patterns of different type of videos, and compare them with the pre-existing results found in the real world social media networks. These patterns show emergent patterns as well, allowing us to comment on various parameters which generated such patterns.

1 Introduction

The emergence of online video sharing sites and social networking have enabled users across the world to create and consume content and videos. However, the sheer volume of available videos makes it difficult for users to decide what to watch. As a result, people have come to rely on their social networks to provide viewing choices. They are more likely to watch videos that are distributed from person to person across social networking sites. How exactly does videos go viral, what are the social contents of the video that go viral, what other factors contribute to the virality of a video? Is any extenal effect playing role in videos becoming viral? The above are some questions which I would explore and address via the model I propose.

The model is an Agent Based Model (ABM). Why is ABM a good choice for modelling virality? There are a number of factors why ABM is the best choice for this model. In ABM aggregate behaviour knowledge is not required apriori. ABM also incorporates probability versus determinism with balance and flexibility. It can generate new hypothesis. Apart from these, ABM is very good at modelling stochastic behaviour and agent environment interactions. In this specific model, the agents are the users and the environment is the web (i.e the videos).

This model has a lot of practical applications and is quite relevant in the real world scenarios. There are millions of people who have jobs directly related to the phenomenon of virality. There are millions of agencies and companies running whose sole purpose is to create viral ads, videos for their customers. People even love watching viral videos, and a lot of companies try to associate themself in order to promote their products. The ad-technique is already a very large industry, and thus knowing the parameters that can affect the virality is becoming a billion dollar business.

This paper follow the following hypothesis **H**: Virality is directly proportional to the sharing likelihood, the network it is used in and social parameters of a video. And participants viewing a video eliciting positive emotion were significantly more likely to forward that video. Other alternate hypothesis can be explored as we go further into the paper.

Finally, the goal of this paper and the model is to recreate the distribution of viewership and their sharing rates of videos based on their type as given in literature (Broxton 2010) and explore what are the parameters that led to the virality.

2 Methods

While no one understands the precise recipe for making a video go viral, many experiments have looked at the characteristics of viral videos as well as the ways in which they were most widely spread.

One such research has been done by a team at Google, Broxton 2010, where they explored on how popularity of a video is affected by social and non-social factors. Broxton defined social factors of a video to be the phenomenon of one person sharing a video to another person in its social graph. A non-social factor is getting recommendation from the platforms algorithms, say YouTube's recommendation.In this model we have considered the social factor which is represented by the social index of the video, as will be explained later. The non-social aspect is addressed by having similar type of videos being close to each other in the environment (patches of the model). I will address these concepts as we move forward in the paper.

All the results and analysis in the Broxton(2010) paper was based on data collected using 1.5 million videos that were randomly selected from the set of videos uploaded to YouTube between April 2009 and March 2010. The data available for each video included the video category (e.g. "Pets", "Music", "News" etc) and the number of views by referrer at the daily level. Here the term 'referrer' is used to describe the person who shared the video to current viewer (Broxton 2010).

Category Name	Highly Shared Video %	Category Social View %	Overall Social View %
Pets & Animals	**42.3**	48.4	1.3
Nonprofits & Activism	38.8	54.7	1.4
News & Politics	31.7	47.2	6.6
Travel & Events	29.5	46.0	1.1
Education	28.8	**61.3**	2.9
Science & Technology	28.4	50.6	2.9
Sports	28.1	39.7	7.9
People & Blogs	26.7	43.8	11.4
Autos & Vehicles	23.8	42.0	2.2
Comedy	20.0	42.9	10.3
Howto & Style	19.7	38.8	2.6
Entertainment	15.6	32.4	16.9
Gadgets & Games	15.3	36.8	6.3
Film & Animation	14.0	33.3	5.0
Music	12.8	29.9	**18.2**
Shows	9.8	34.1	3.0

Figure 1: The sharing % of 16 different types of videos over 30 days (Fig 5: Broxton 2010)

Figure 1 summarizes the results of sharing percentage of different types of videos over 30 days from YouTube. From this data, I worked upon the "Highly Shared Video %" category. This provided the data for sharing likelihood of 16 different types of videos over 30 days.

Guadagno explored various hypothesis about emotions effect on the virality of a video. He researched if certain type of video has more chance of being shared or not. For instance, one of his hypothesis was that video having higher social index in terms of video being funny or cute or happy is more likely to be shared, however another hypothesis was that videos with anger, sadness and disgust were more likely to be shared. He conducted several experiments and did not came to a proper conclusion to reject any of the above hypothesises. So I expressed the emotions in terms of social motivations proposed by Nottingham (WistiaFest, 2015). The social motivation types are as follows:

Zeitgeist, Opinion-seeking, Experience, Perseverance, Kudos,Social-welfare , Reaction, Expression, Utility. The social motivation index for any video is calculated by normalizing the sum of all the indexes, unlike Guadagno, who used notion of happiness, sadness etc. This will server as a apt substitute for video emotion. The higher the index more likely it will be that the video will be shared. This index will be a representation of how the social motivation of a video, depending on the mood of a person, will affect their chances of sharing or not.

Apart from emotions of videos, Guadagno also conducted experiments to see if the shared content depends on the type of person who shared that content (influencer, friend etc), but was unable to find any concrete evidence for any. Therefore, there are no groups as such in the model.

The model contains users in the social network, who are represented by agents. The user tries to view video at the location he is on (the video on the patch the user is), if it has not already viewed the same video or randomly watches the video again with a probabilty of 50%. Now, if the user watches the video, then it decides if it wants to share this video or not.Now for deciding if the user wants to share the video or not, the user evaluates the mean of the social-indexes of the video, and multiplies it with his likelihood of sharing the video, and the likelihood of this video being shared and shares it with a random probabilty. If this video is shared, then the connection strength between the user who suggested this video to this user(previous-recommender) is increased by 1.

The model contains the videos in form of patches. There are 16 video types as mentioned in the figure 1 (Broxton 2010). The types are : type 1: Pets and Animals, type 2: Nonprofits & Activism, type 3: News & politics, type 4: Travel & Events, type 5: Education, type 6: Science and Technology, type 7: Sports, type 8: People & Blogs, type 9: Autos & Vehicles, type 10: Comedy, type 11: Howto & style, type 12: Entertainment, type 13: Gadgets & Games, type 14 Film & Animation, type 15: Music, type 16: Shows

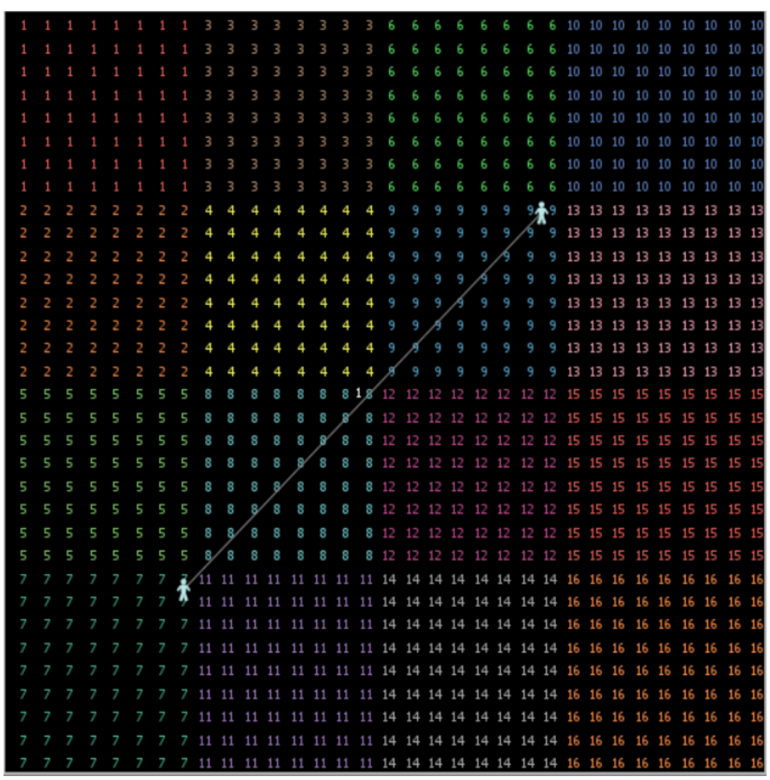


Figure 2: World divided into 16 parts showing the video type of each of the 16 sections

Figure 2 shows the world is divided into 16 parts and the video types related to each other are placed in proximity of each other. For example, all the videos of type 1 are together. The video-sharing-likelihood is set to [42.3 38.8 31.7 29.5 28.8 28.4 28.1 26.7 23.8 20.0 19.7 15.6 15.3 14.0 12.8 9.8] (Broxton 2010 Figure 5). The patches are being used as videos. In order to identify each video uniquely, I assigned a unique number to each video, called the video-id, so that the video can be identified uniquely.

There is a social network in the model. The models are of three type:

- Preferential : Users are more likely to be linked with people who have several links in a preferential network
- Small World : A user is no more than 6 degrees of separation away from another person.
- Random : Users are connected to each other with a probability of 0.1

A user is connected to other turtles via undirected links, with weights on the links. I have used these links as a means to represent connection strength and trust on social media. Trust between users increases as they share content between each other, and the more the content sharing, the better is the trust (Guadagno 2013).

To watch a new video, a user checks if there are any connections of it who are trying to share/recommend some video. If so, out of of all those users, it selects the user which it trusts the most (here trust is represented by connection(link) strength). Upon finding such a user, this user moves to the location where the video recommended by the sharer user is (this process is done by teleporting this user to the patch location of the sharer user). And the previous-recommender of the user who teleported is set to the user who shared the video.

The more you like the videos that someone sends you, the stronger your connection becomes. This simulates the likelihood that you will watch a video sent to you from someone who has a significant influence over you, such as a personal friend or a celebrity. The more videos you post and like for each other, the more powerful the person's presence becomes.

In the real world also, the videos are removed after some time, or even if they are not removed, then the number of views garnished by them are minimal. Therefore, after some regular interval of time (decided by the slider video-removal-rate in the model) there is a chance that a video will be removed, and in place of deleted video, a new video will be added of the same type. The chance of deletion is also decided by a slider in the model called video-removal-probablity.

3 Results

The parameters have been calibrated so that 100 ticks represents a day and the model runs for 1 month as in Broxton 2010 paper. Also, the total number of videos taken were 1.5 million, and using parameter calibration, I was able to fix the total number of videos to be around 1.5 millon at the configuration in Figure 3. Figure 3 shows the complete model with the interface and the results for 250 number of users, with preferential network, video removal time of 3 days (300 ticks) with a probability of 0.05. The model shows the visual output of the simulation along with the sharing rates and number of views distribution over different type of videos.

Figure 4 shows the comparision between the average number of views of different types of videos from the model and the Broxton paper. It is evident from this comparision that our hypothesis was correct and the number of views are directly proportional to the sharing rates and sharing likelihood. This is an emergent pattern which was not coded into the model, but is an outcome of the interactions between agents and environment.

Figure 5 shows the sharing rates of different types of videos over 30 days. These are the emergent results of the simulation. From the figure 1 we can see the sharing rates that were formulated by Broxton (2010). The sharing rates graph and the sharing percent from the table are nearly the same. For instance, consider the video type of pets and animals. The sharing rate in the table is 42.3 % where as in the graph, it is also around 42 %.The Nonprofit videos had a sharing rate of 38.8% in the table, where as in the simulation graph is also around 30% and so on. Again, these information were not coded into the model, but are the emergent outcomes of the model. Figure 6 shows the sharing

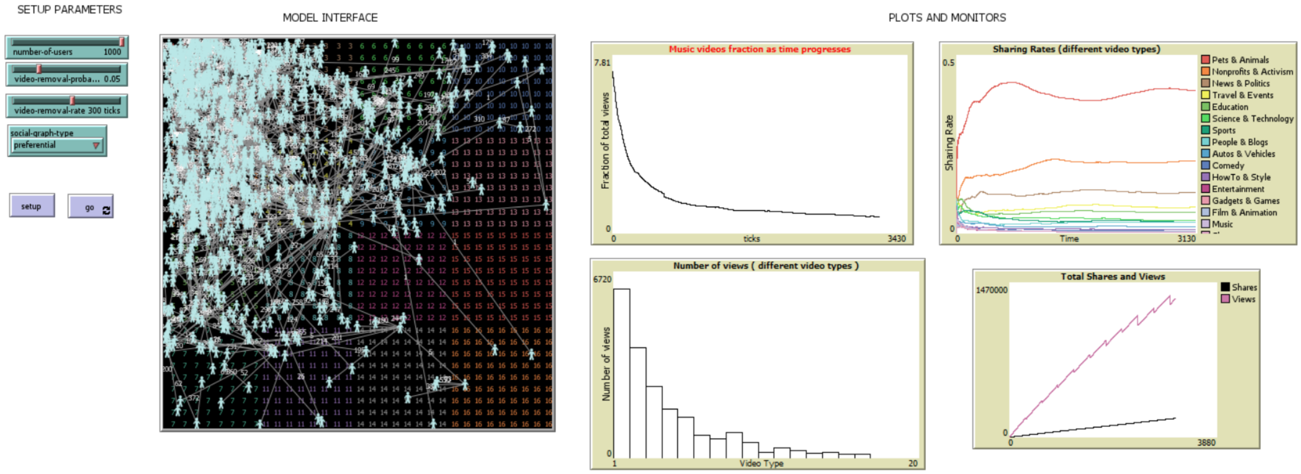


Figure 3: Overall model

rates of just the music video. And we can compare from the rate given in the Broxton's paper, that they are again the same.

Figure 7 and 8 shows the results of simulation on small world model. The model interface given a correct visual description of the simulation. It shows that most of the users will be accumulated around very small set of videos of same type. This is because in a small world model, almost all the users are connected to each other, and no user is more than 6 connection hops away from another user. For this reason, this simulates the case when all the users are of very similar interests. The sharing rates does not match with the Broxton's analysis, and are very off. Also, from the viewership plot, it is easy to see that 90% of the views are for the video type 1, which has the most sharing likelihood.

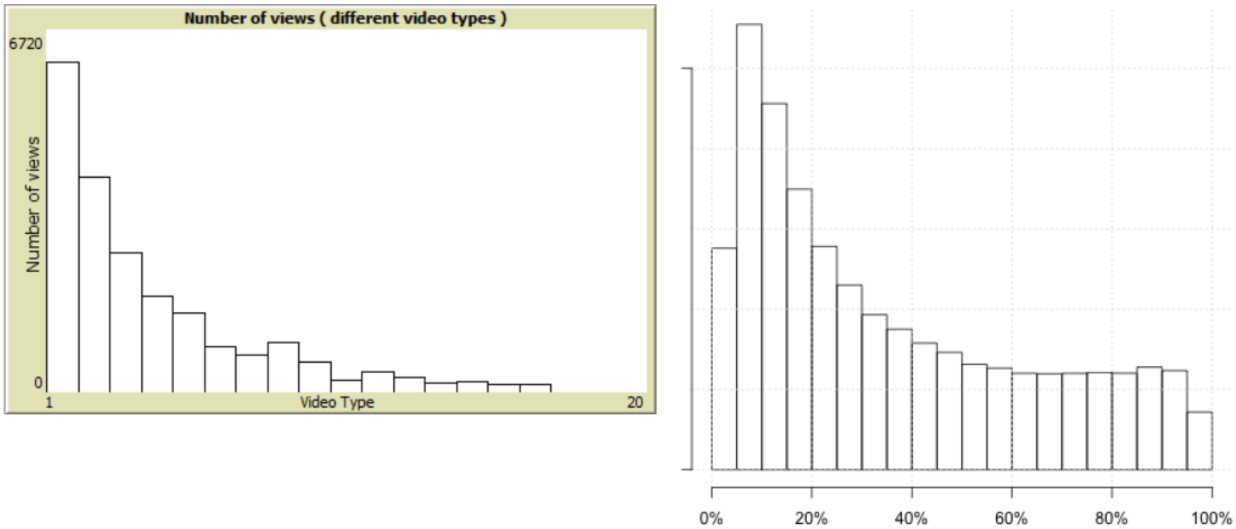
Figure 9 shows the sharing rate and average views for random graph. The graph is formed with 10% chance of any 2 users being connected. And the results are even worse than that of the small world model. They are completely random.

Figure 10 is the configuration where the video deletion rate was set to 12 hrs only (50 ticks) in a preferential network. And the probability of deletion was set to 20%. It is clearly visible that the number of views becomes almost constant when the videos are removed too frequently. The number of views are removed for the deleted video. This result also matches the expectation of the real world. If the videos would have been removed too frequently, then there would have had not been much about virality. However, there is not much effect in the distribution of views or the sharing rates. They are the same as expected.

4 Discussion and Conclusion

It is quite evident from the above analysis that the preferential model is the one that best replicates the real world social network. The results of preferential model were the closest to Broxton's analysis where as other networks were not even close. While other attributes such as video-deletion-rate, deletion-probability etc were also tested in Behavior Space, the network-type had by far the greatest effect on the results. From these observation we can conclude that although the quality of the video is significant to its virality, the platforms in which it is transmitted have a greater effect. If the content of a video resonates with one user in a preferred network, it is more likely to be shared with someone who might enjoy it. This discovery may be crucial for anyone looking to make a viral video in the future, and this model will be highly relevant in such cases.

There are couple of things which can be implemented in future. First, there is no homophily. For example, users with common preferences can be linked and have stronger ties between them. Similarly for videos with similar social indexes can be grouped together. Second, there can be some celebrity



(a) Simulation Plot for mean number of views for each type of video

(b) Broxton's analysis for viewing distribution over 30 days

Figure 4: Comparison between Broxton's analysis and Model simulation analysis

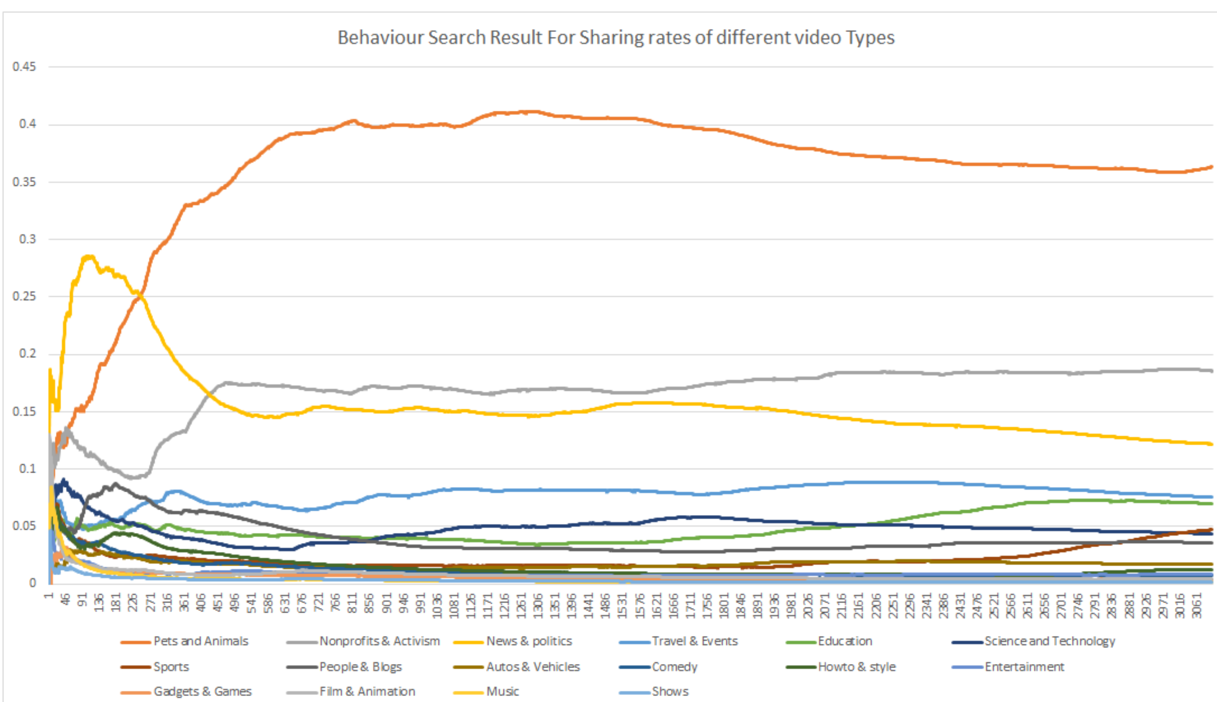
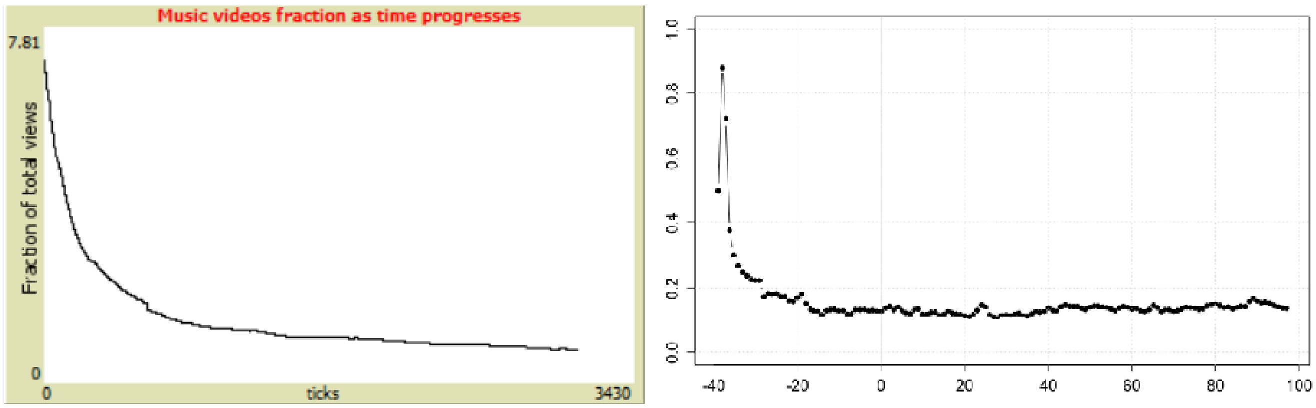


Figure 5: Sharing rates of different types of videos over 30 days



(a) Music Video sharing rates

(b) Broxton's Music video sharing fraction over time

Figure 6: Rate of sharing for just the video type music

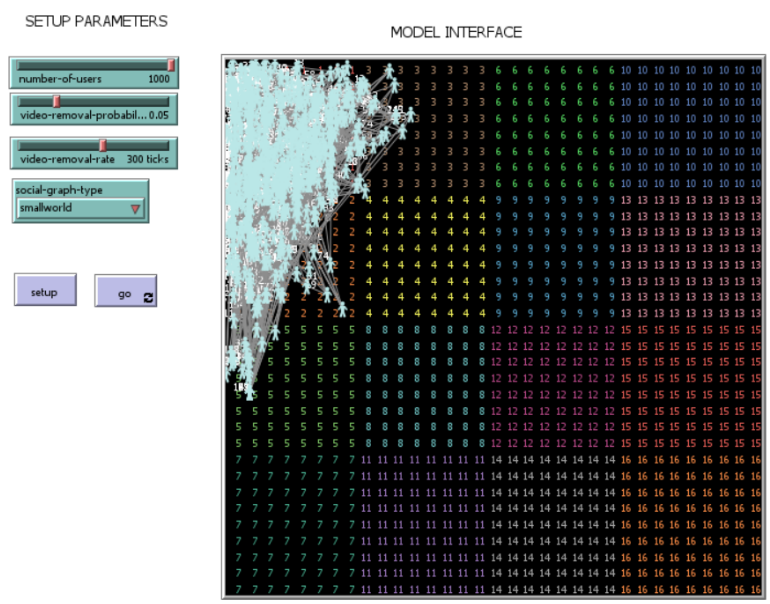


Figure 7: Small World model Interface

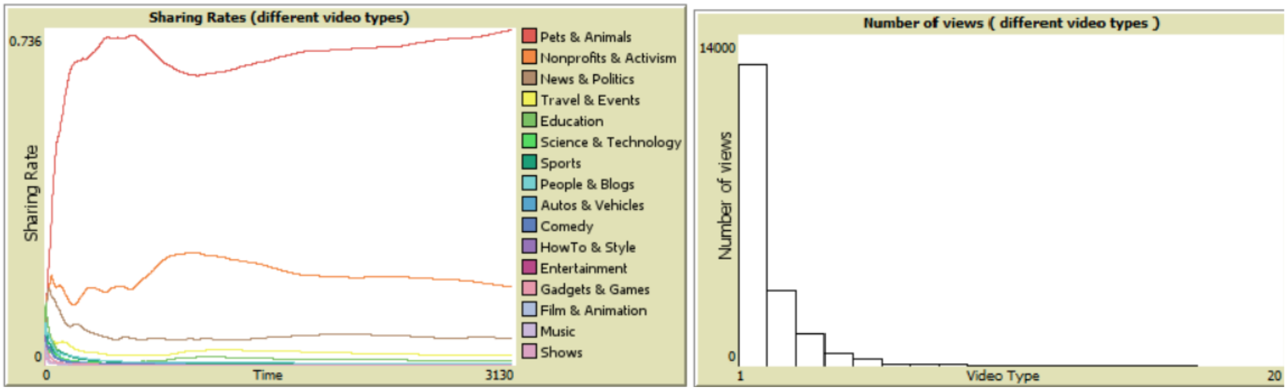


Figure 8: Small world model Sharing rates and mean number of views for different types of videos

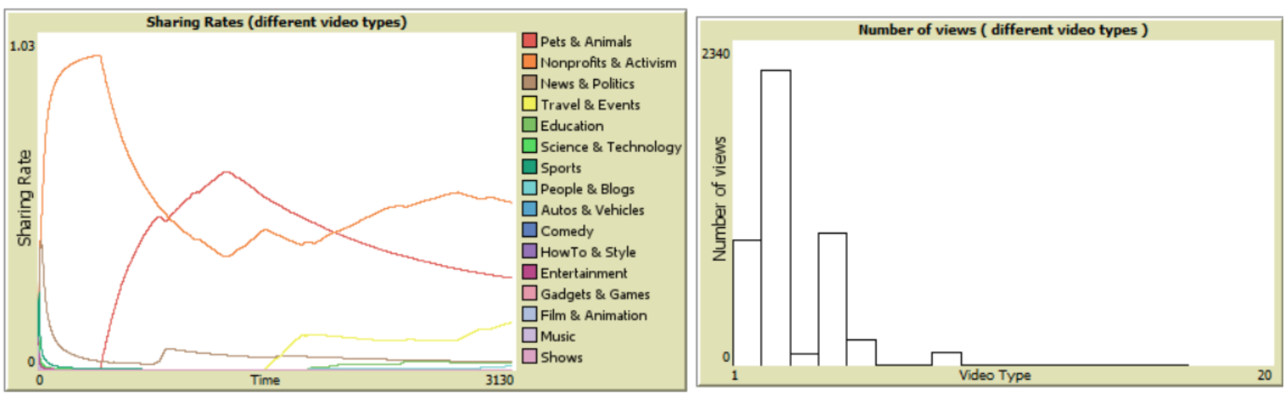


Figure 9: Random model Sharing rates and mean number of view for different types of videos

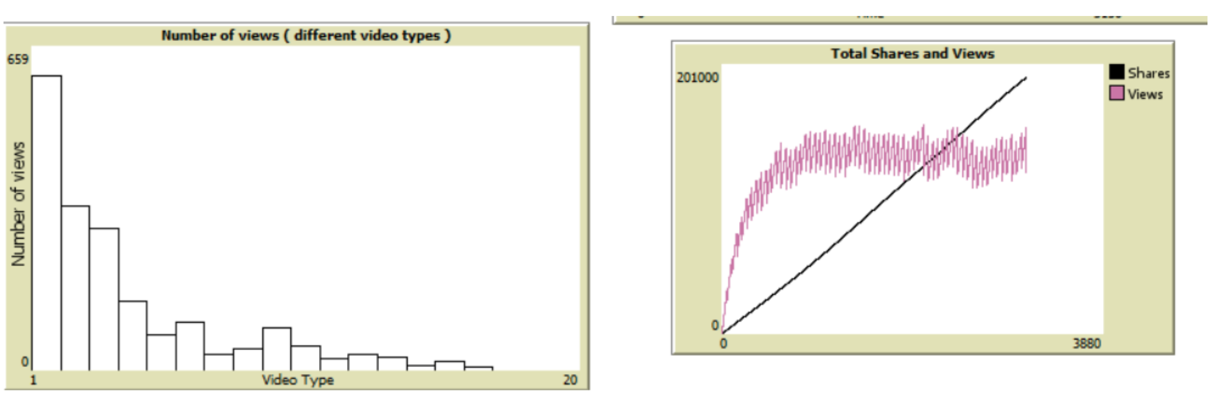


Figure 10: Number of views and total shares for deletion rate of 1/2 day and probability 20%

users who can have much more influence than normal people, and their connection networks can be huge The popularity of similar videos can be influenced by this person's influence. Third, the model can be very easily extended to predict the virality of contents other than videos, such as memes. These concepts can be implemented in future.

There are already a lot of alternate hypothesis in the field of virality of content. Our hypothesis **H:** "Virality is directly proportional to the sharing likelihood, the network it is used in and social parameters of a video. And participants viewing a video eliciting positive emotion were significantly more likely to forward that video" is not the only hypothesis that may be correct. For instance consider the alternate hypothesis $\mathbf{H_1}$**:** The virality is not directly proportional to the sharing likelihood, but depends on a combination of social parameters. Broxton's analysis indeed shows that there are chances of accepting the alternate hypothesis, as in Figure 1, the over all social view % is much higher for music videos instead of pets and animals videos. There is no denying the results formulated by research. But still there is not enough evidence to reject out hypothesis **H**. Another alternate hypothesis might be that the video having higher social index in terms of video being funny or cute or happy is more likely to be shared. We can test our hypothesis against all sort of alternate hypothesis.

Data also suggests that back in 2012, during the US presidential elections, a large amount of credit of obama's victory goes to the viral campaign videos that were published on youtube. Millions of people watched and shared these videos. To conclude and shine some light on the relevance and motivation of this model, there might never be a recipe for making the perfect parameters for viral video, a growing number of businesses are getting closer to it. Advertising companies are studying behaviours of users sharing the video as well as the features of the video itself to see what leads to its virality, and they appear to be able to determine when a video would go viral with an accuracy of nearly 80%. Unlocking the secret of viral videos to build effective viral marketing campaigns may be a game-changer in the ever-changing billion-dollar advertisement market, so studying the actions of this and other related models could be a worthwhile investment.

References

[1] Tom Broxton, Yannet Interian, Jon Vaver, and Mirjam Wattenhofer. Catching a viral video. In ICDM'10 Workshop, pages 241–259, Los Alamitos, CA, USA, 2010. IEEE Computer Society

[2] Phil Nottingham. Building a Social Video Strategy- WistiaFest 2015

[3] Guadagno, Rosanna Rempala, Daniel Murphy, Shannon Okdie, Bradley. (2013). What makes a video go viral? An analysis of emotional contagion and Internet memes. Computers in Human Behavior. 29. 2312-2319. 10.1016/j.chb.2013.04.016.