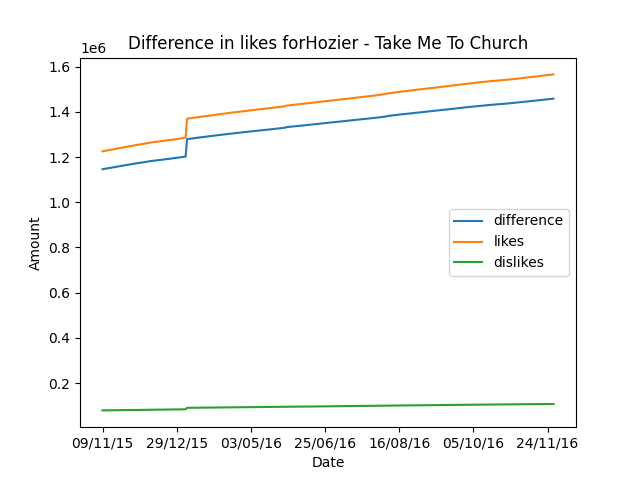
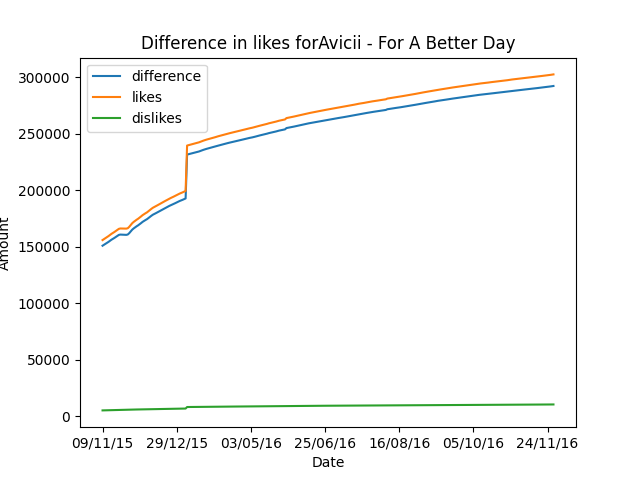
Report week3&4

Jelle van den Brink - s2743450

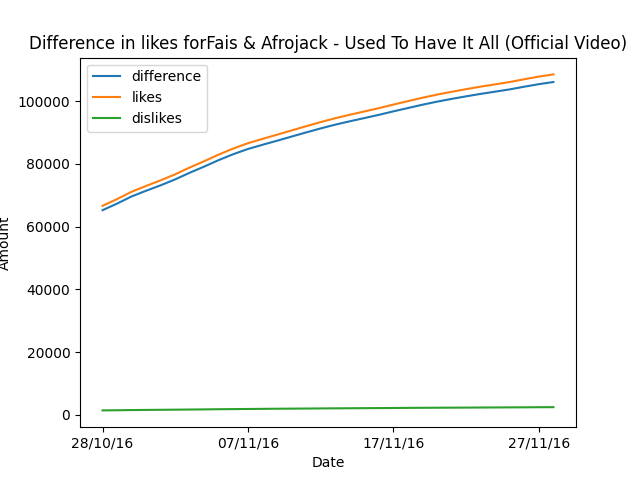
Tim Angevare - s2744007

## Cascading effects

1. These graphs all follow the more or less the same trend. This starts with a linear fairly steep slope for the first part of the graph. This then turns into a sharp increase for a small period of time (that we believe is the cascade), which again turns into a more or less linear trend with a decreasing slope.

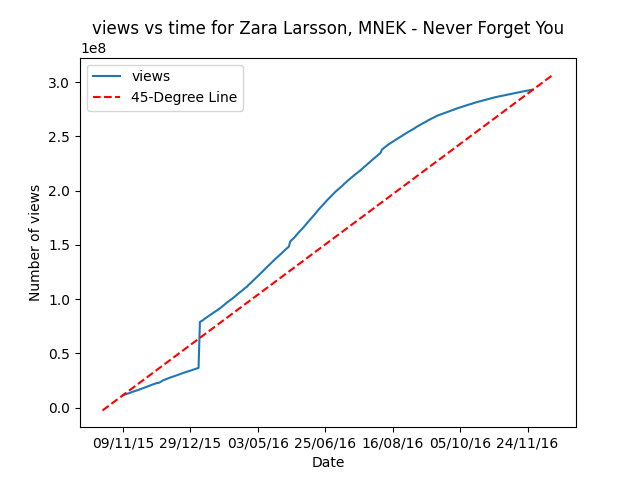


1. The differences of likes and dislikes plotted for the non top 100 songs seem a bit different. These graphs also more or less follow the same pattern, namely this is a positive slope following a more or less linear trend and decrease in slope towards the end of the graph. We believe that these do not contain cascades of any kind.

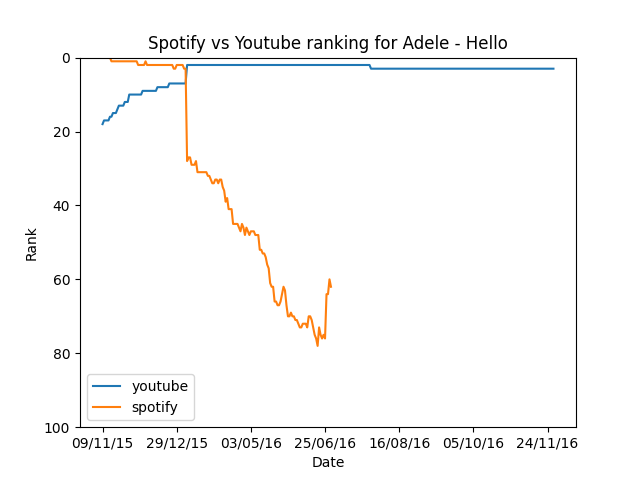
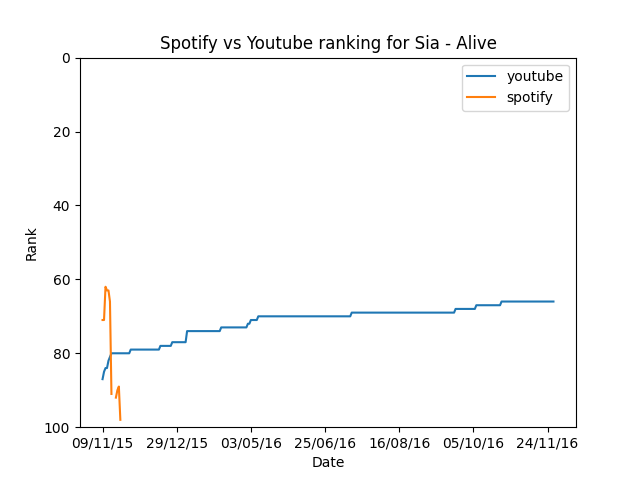


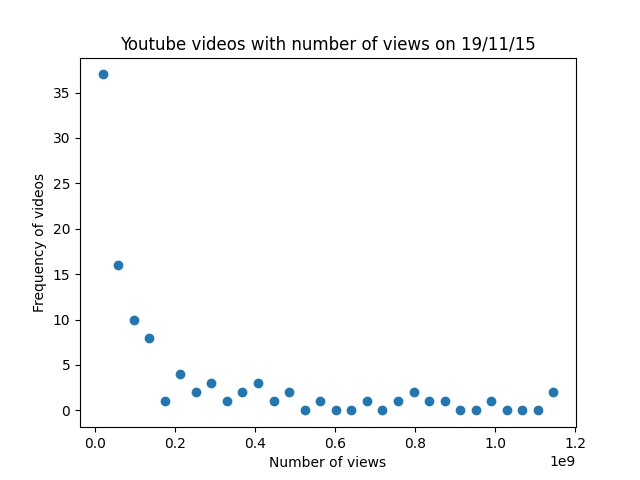
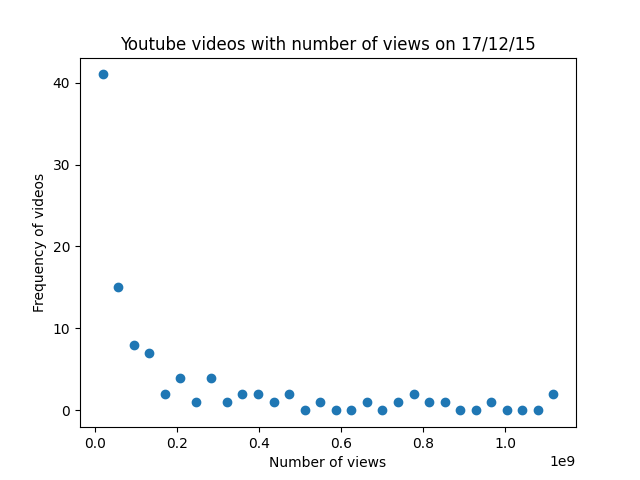
1. If we compare the two different graph patterns we can see some interesting observations. Namely, the graphs of the songs part of the top 100 chart all contain a sharp peak while the ones that do not get into the top 100 do not contain these. Furthermore, for the songs in the top 100 after the sharp peak, the line will follow more or less the same trend as the not top 100 songs. This leads us to the conclusion that in the top 100 there is cascading. The songs start by being recognized for their improved quality (the steeper linear slope at the start) then they get picked for the top 100 and the cascading starts since the song becomes famous (the steep spike). And then the initial fame dies down as it follows a normal trajectory (the same sort of pattern as in the non top 100 songs).
2. Seeing the plot I believe that the likes / dislikes represent the high / low signals. Other factors that can contribute to the cascading effect could for example be the amount of views on a video or the subscribers to a channel. Linking this back to the lectures one can see the previous two metrics as the amount of people standing in a line to watch a video (subscribers) or a counter of the amount of people that have stood in the line before you (views on a video)

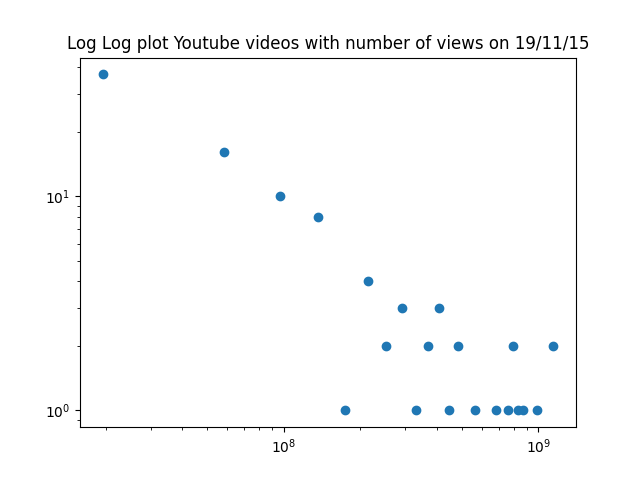
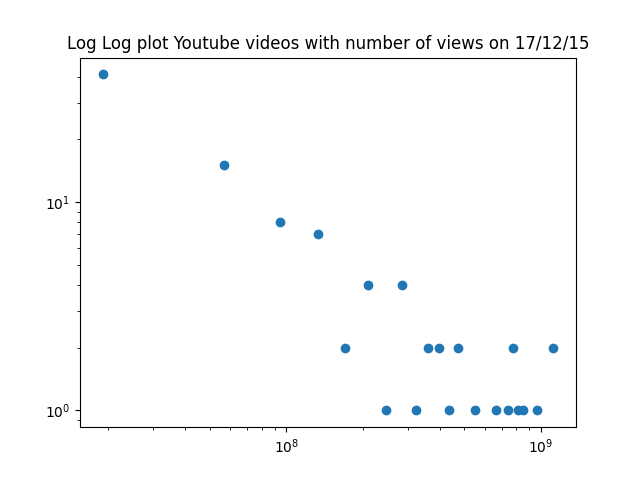
## Network effects

1. What we see in the view plot of the youtube videos is that there usually are equilibrium values among the start of the plot. Further when we increase, the plot line does not seem to intersect with the 45 degree line anymore. A shortcoming we saw here is that most of the videos are not published at the same time the dataset starts, so the view count does not start at 0. Because of this the 45 degree line also does not start at 0 and the equilibrium might not occur at the correct values. Nevertheless for the song “Never Forget You” the theory was proven as there were multiple equilibrium values and in the end, the plot converges towards the last equilibrium.
2. The phenomena of music preferences coincide quite closely with the typical cases the network effects model is intended to describe. If a video on Youtube is watched a lot it shows up in the trending category and so it is recommended to more people to view it, this is a strong network effect. The same happens with person-to-person communication, because when people like a song they also recommend it to other people thus showing the network effect.
3. For the different variables there can be multiple metrics to represent these. The combination of multiple would be the most accurate. Some metrics are present in the dataset other are not.
   1. Intrinsic Interest r(x) can be the combination of a lot of values such as the length of a video, the quality, relevance to the viewer, title, reputation of the creator
   2. Network effects f(z) can be the amount of comments as well as the amount of shares for example.
   3. Price can be the revenue made from the video from a creator’s perspective as well as views from a visitors perspective

## Rich-get-richer popularity effects

1. There are power laws in effect here as there is a line present in the plot. Only around the x>10^8 and y<10^0.5 this line is gone and the data is scattered. The power laws are therefore weak and there are almost no get richer effects.
2. If the views of the next day are proportional to the views of the last day this implies exponential growth. Since the video has a proportional increase in growth the potential viewers will also increase, which in turn increases the amount of views leading to exponential growth.
3. Studying the amount of views plotted against time in assignment 2, we cannot see any exponential growth. The trendline here is for the most part linear. 
4. In the plots below some graphs comparing the Spotify vs Youtube rank can be seen. The outcomes are not always in line, as the songs in the youtube playlist are constant over time, but these songs are not always in the spotify top 100 anymore after a year. For Hello by Adele it can be seen that it stays in the Spotify top 100 for quite some time. However for Alive by Sia it can be seen that its rank in the youtube playlist increases slightly but for spotify it is out of the top 100 quite fast.



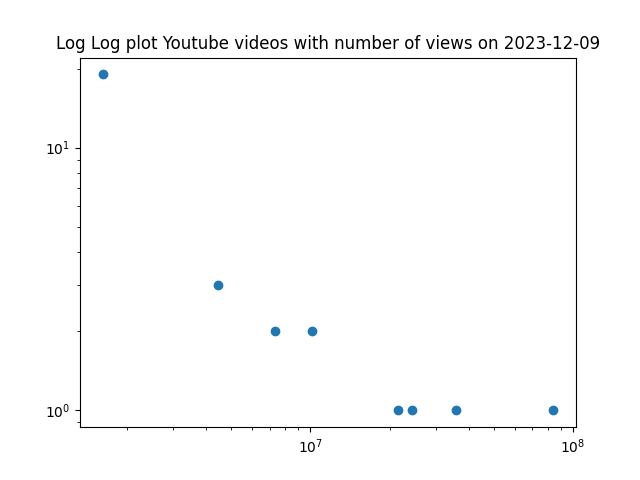
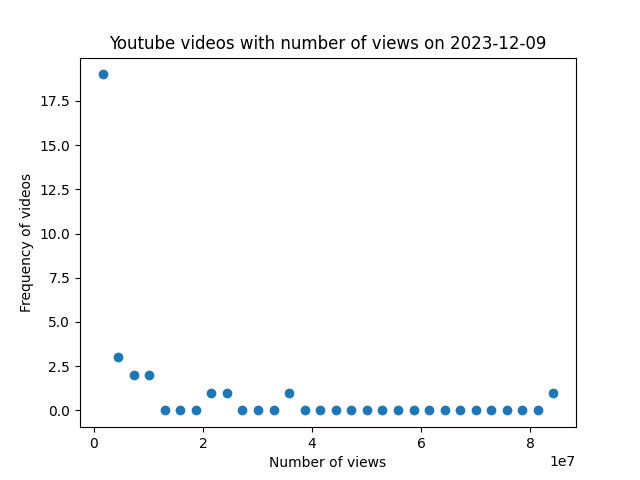


## Information diffusion

1. The model of information diffusion does apply to music preferences a bit, but not very much. The concept of people being very open to all kinds of music as long as other people also listen to it does exist, but most people stick very much to their own genre of music, no matter how may of their friends listen to it. This is because a big part of people listen to music on their own and do not gain a lot from friends listening to the same music, but this obviously does not apply to everyone.
2. The model of information diffusion is kind of related to the rich-get-richer phenomenon, as they are both based on people making decisions based on other people. But the eventual core is different so that is not related.

## Create your own dataset

1. See jupyter notebooks
2. As can be seen from the graphs below, there is a long tail visible, but the tail goes down so quickly that it is not very clear, apart from the one dot in the top left corner. In the log log plot this effect does become more clear.



## Conclusions

1. Based on the available datasets it can be concluded that of the four effects the model of information diffusion probably fits the worst to the data, as people relatively do not get a lot of gain from listening to the same music as their friends. The cascade effect is also not very clear, even though all of the graphs show one steep increase in views, because of the date this happens this is probably related to something else. For all graphs the line gradually increases in a similar trend, so there is no strong cascading effect that shows up. Two effects that were more obvious were the effects shown in the network effects section amd the rich-get-richer section. As can be seen in the figure in the network effects section, the song Never Forget You by Zara Larsson has a clear graph in which network effects are visible. This can also be said for the graphs shown in the rich-get-richer section, even though they are less clear. There is however a flattening curve that can be seen, yet it does have a few outliers. To conclude, the network effect best explains the data in the set.
2. To correctly investigate each of the phenomena the most important data would be the streams, obviously gathered for a period of time. For youtube this would mean video views, and for Spotify the streams. That is the most important, but a bit more data can be added to gain some interesting other insights. For example likes/dislikes, but also data from other sources like the popularity of artists, social media posts, big events can all have an influence on these phenomena and could be interesting to explore.