An Analysis of YouTube Trending Videos with GCNs

An ORIE 4741 Project Proposal

Lorenzo Vettimo (ls769) and Jeremy Wang (jw2363)

Fall 2021

YouTube Data Analytics
Cornell University
Ithaca, NY 14850

Dear YouTube Data Analytics Team,

As students of Cornell University, we would like to thank you for serving as a large entertainment platform for billions of users across the world. With the rise of numerous trending videos from various topics like Sports and Make-up, YouTube has a truly broad range to offer in its services. After spending numerous hours accumulating to what may be many months of our long lifetimes, we propose performing a YouTube Trending Video dataset analysis.

The application of being able to gain insights into such a dataset would allow for increase user retention and thus environment by others to recommend similar videos. Based off of previously watch YouTube videos in which a person liked or disliked, we wish to determine if they may like the video based on fundamental characteristics like the "Title" or "Video Description" where text will be used as keywords so that we can use counters instead of more involved NLP algorithms. The dataset we have is the YouTube Trending Video Dataset (updated daily) on Kaggle. The innovation we bring in comparison to existing studies on YouTube analytics is the use of a graph neural network (GNN). Specifically, we propose the use of a graph convolutional neural network (GCN) in order to learn features by inspecting neighboring nodes. The use of graphs as a way to illustrate and understand relationships in data is essential for the task of video analytics due to the numerous and complicated dimensions the data consists of since there are intricate relationships that the features have with one another. In addition, to our knowledge, an investigation of YouTube analytics using a GCN has not been published, so we strongly believe such an exploration could lead to insights for the team. With the success of GCNs in tasks like drug discovery and combinatorial optimization, we believe there may be new insights that can be gained from our study.

If you consider approving our project proposal, however, we must mention that in conjunction they come with additional limitations. This is because of the challenge of making our algorithms work well both individually and as a cohort in terms of what is available within the dataset. Therefore, we leave the option for you to elect aspects of our solution that may be used to drive insights instead of the whole package. We always seek to help creators and users in order to ultimately inspire new creative videos, especially with the rise of shorter and fast-paced videos on TikTok. We hope you take our project into consideration and reach out to us if you have any comments or concerns about our proposal.

Sincerely,
Lorenzo Scotto di Vettimo and Jeremy Wang