Beyond the Views: The Real Impact of YouTube Popularity on Creators and Society

Roman Shumliakivskyi
r.shumliakivskyi@studenti.unipi.it
Student ID: 683962

ABSTRACT

Every creator aspires to make a significant impact and achieve widespread recognition. But what factors truly drive a creator's success on YouTube? Is it purely the content's quality, or do collaborations and connections with other popular creators play a pivotal role? This paper, titled "Beyond the Views: The Real Impact of YouTube Popularity on Creators and Society," aims to analyze whether a YouTuber's popularity is linked to their collaborations and connections within the platform, using network analysis.

The study was conducted on a network of over 10k data points, comprising connections between channels (edges) and individual channels (nodes). Each edge represents a video collaboration, including details such as the source channel name, target channel name, video title, and view count. Nodes contain information on channel IDs, genres, view counts, and labels. The data was collected using the YouTube Data API v3 and analyzed using the Python library NetworkX.

Our analysis aims to uncover the formula for success for video bloggers by examining the relationship between the number of views and the extent of their network connections. Initial findings suggest that a YouTuber's popularity is often enhanced by collaborations with other popular creators. However, exceptions exist where some creators achieve high popularity independently, while others with numerous connections remain less prominent if their collaborators are not widely recognized.

1. INTRODUCTION

YouTube Data API v3 provides a comprehensive toolset for accessing a wealth of information about channels, videos, and interactions on the platform. It enables developers and researchers to collect data on various aspects of YouTube activity, making it an invaluable resource for analyzing trends and patterns within the platform. This paper leverages the YouTube Data API v3 to investigate the factors contributing to a YouTuber's success.

Every creator aspires to make a significant impact and achieve widespread recognition. However, understanding the dynamics behind a creator's success on YouTube involves examining multiple factors. While the quality of content is undeniably important, this paper explores the hypothesis that collaborations and connections with other popular creators play a crucial role in enhancing a YouTuber's popularity.

The aim of this paper, titled "Beyond the Views: The Real Impact of YouTube Popularity on Creators and Society," is to analyze whether a YouTuber's popularity is linked to their collaborations and connections within the platform, using network analysis. The study focuses on constructing and analyzing a network of over 11,315 data points, representing connections between channels (edges) and individual channels (nodes).

Each edge in the network corresponds to a video collaboration, including details such as the source channel name, target channel name, video title, and view count. The nodes, representing individual channels, contain information on channel IDs, genres, view counts, and labels. By utilizing the Python library NetworkX, we conducted a thorough analysis of the data collected using the YouTube Data API v3.

Our analysis aims to uncover the formula for success for video bloggers by examining the relationship between the number of views and the extent of their network connections. Initial findings suggest that a YouTuber's popularity is often enhanced by collaborations with other popular creators. These collaborations can lead to increased visibility, shared audiences, and cross-promotion, which collectively contribute to higher view counts and subscriber growth.

However, exceptions exist where some creators achieve high popularity independently, relying solely on the strength of their content and unique appeal. Conversely, there are cases where creators with numerous connections remain less prominent if their collaborators are not widely recognized or fail to engage a broad audience. This study seeks to provide a nuanced understanding of the interplay between content quality, collaborations, and network connections in determining a YouTuber's success.

By shedding light on these dynamics, this paper aims to contribute to the broader discourse on digital influence and the factors that drive online success. The insights gained from this study can be valuable for aspiring creators, marketers, and researchers interested in the evolving landscape of social media and digital content creation.

2. DATA COLLECTION

The data collection process, hinged on the strategic use of the **YouTube Data API v3**. This powerful interface provided the means to access a wealth of information essential for constructing a comprehensive network of YouTube creators, with a primary focus on video collaborations, view counts, and channel-specific details.

Authentication and authorization were paramount. **OAuth 2.0**, a well-established protocol, was employed

to ensure secure access to the API. Permissions were meticulously defined, using the **`youtube.readonly`** scope, limiting actions to data retrieval and safeguarding against unintended modifications. A **`client_secret.json`** file, containing sensitive credentials like the client ID and client secret, was securely stored and managed to facilitate this authentication process.

API requests, carefully crafted and optimized, served as the primary tools for data retrieval. These requests targeted two main types of information:

Video Collaborations: Extracting the source channel, target channel, video title, view count, and publication date of collaborative videos provided the foundational links for the network analysis.

Channel Information: Gathering details such as channel IDs, genres (where available), total view counts, subscriber counts, and relevant labels or keywords enriched the dataset with contextual information about each creator.

The retrieved data, however, was not without its imperfections. Duplicate entries, missing values, and inconsistencies required meticulous cleaning and preprocessing. This ensured data quality and reliability, paving the way for subsequent analysis.

Structured formats, **primarily CSV files**, were chosen for data storage due to their versatility and compatibility with a wide range of data analysis tools and techniques.

Throughout the data collection process, challenges arose that demanded creative solutions:

API Quotas: YouTube enforces usage limits to ensure fair access. Request optimization and, when necessary, a Google Cloud Platform subscription with increased quotas, addressed this constraint.

Rate Limiting: To comply with the API's rate limits, requests were thoughtfully spaced out, preventing disruptions due to excessive usage.

Data Volume: Popular channels generate vast amounts of data. Efficient storage and processing strategies were employed to handle this volume effectively.

Data Output:

Upon completion of data collection, we compiled a dataset comprising 30 YT channels connected by 11,316 edges, adhering to our connectivity assumptions. Alongside videos connections, we gathered supplementary metadata such as views.

3. NETWORK ANALYSIS

We analyzed our network of artist collaborations and compared it to networks generated by the Barabási-Albert (BA) and Erdős-Rényi (ER) models.

A Network Overview

Metric	Value
Avg. Degree	2.12
Network Diameter	26
Graph Density	0.00018411
Connected Components	1722
Avg. Clustering Coefficient	0.146
Number of Triangles	7704
Avg. Path Length	9.37

Our network has an average degree of 2.12, indicating it is relatively sparse. The network's diameter is 26, meaning the longest shortest path between any two artists is 26 steps. The density of the network is 0.00018411, reinforcing the sparsity. There are 1722 connected components, indicating a high level of fragmentation. The average clustering coefficient is 0.146, showing moderate local clustering, and there are 7704 triangles, supporting the presence of locally connected groups. The average path length is 9.37, indicating a moderate level of connectivity.

BA Network Overview

Metric	Value
Avg. Degree	1.9998
Density	0.000176
Avg. Clustering Coefficient	0.0
Number of Triangles	0
Connected Components	1

The BA model (figure 1) produces a network with an average degree of 1.9998 and a density of 0.000176, both slightly lower than our network. The BA model has no clustering (average clustering coefficient is 0.0) and no triangles. The number of connected components is 1, indicating the BA network is a single, large connected component.

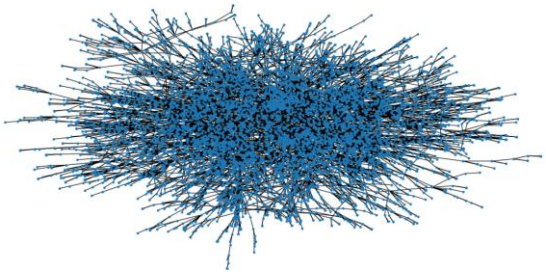

Figure 1: BA Model

ER Network Overview

Metric	Value
Avg. Degree	1.9838
Density	0.000175
Avg. Clustering Coefficient	0.0
Number of Triangles	0
Connected Components	1910

The ER model (figure 2) produces a network with an average degree of 1.9838 and a density of 0.000175, both slightly lower than our network. The ER model also has no clustering (average clustering coefficient is 0.0) and no triangles. The number of connected components is 1910, indicating a more fragmented network compared to our network.

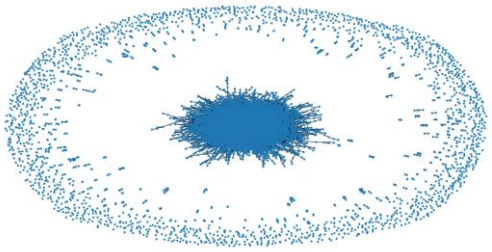

Figure 2: ER Model

The network has a diameter of 2, indicating a high level of interconnectivity as any two nodes can be connected through at most two steps.

The average path length is approximately 1.996, suggesting that on average, any two nodes are nearly directly connected.

The average clustering coefficient is 0.0, which shows that there is no tendency for nodes to form tightly knit groups.

The density of the network is approximately 0.0001757, indicating a sparse network with many possible connections not realized.

Centrality Measures:

Top 5 Nodes by Degree Centrality and by Closeness Centrality:

1. UCJ5v_MCY6GNUBTO8-D3XoAg	0.0496
2. UCPgMAS8woHJ_o_OZdTR7kcQ	0.0491
3. UC2tsySbe9TNrI-xh2lximHA	0.0475
4. UCq-Fj5jknLsUf-MWSy4_brA	0.0470
5. UC-lHJZR3Gqxm24_Vd_AJ5Yw	0.0454

Degree Distribution:

The degree distribution plot reveals that most nodes have very few connections, while a few nodes (hubs) have many connections, characteristic of a scale-free network.

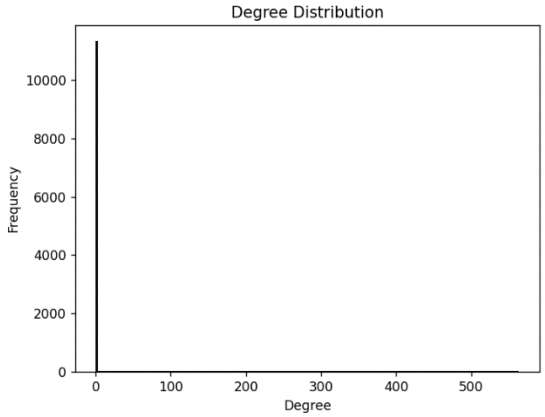

Figure 3. Graph of Degree Distribution

Connected Component Size Distribution:

The connected component size distribution plot indicates the presence of many small components and a few larger ones, typical of networks formed on social media platforms where a few users are highly connected while many have few connections.

This analysis reveals a sparse, highly interconnected network with prominent hubs, typical of social media platforms.

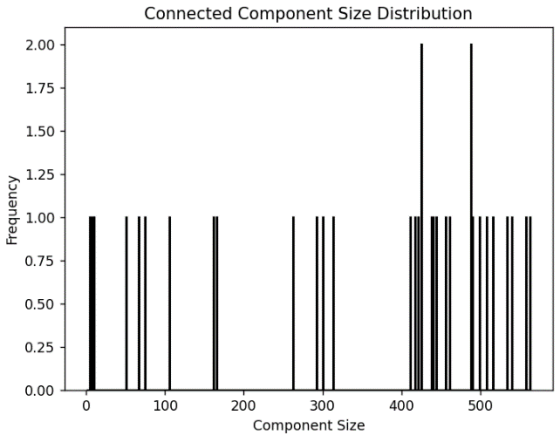

Figure 4. Graph of Connected Component Size Distribution

The high interconnectivity and presence of influential hubs suggest that information can spread rapidly across the network. However, the lack of clustering and the fragmented component structure may limit the depth and reach of this spread within local clusters.

The sparse nature of the network highlights potential opportunities for increasing connectivity and collaboration among nodes to enhance network robustness and resilience.

Overall, the analysis provides a comprehensive understanding of the YouTube network's structure, revealing a highly interconnected yet sparse network with key influential hubs, typical of large-scale social networks.

4. OPEN QUESTION

So far, we have analyzed a network of YouTube creators connected by video collaborations. In this section, we will use the results of the analysis to find an answer to the following question: Does the frequency of collaborations on YouTube correlate with a channel's popularity, and does it impact their overall success?

To answer this question, we considered the view count and subscriber data of the channels involved. Popularity is defined as the number of views and subscribers a channel has. We will identify the nodes with the highest degrees (indicating numerous collaborations) and compare their view and subscriber counts to determine if more collaborations lead to greater popularity.

First, we re-drawed the network, focusing on the central component where there are the most nodes and the most connections. In the visual representation, we will use a heatmap-style color palette to represent the popularity of the channels: the darker the color, the higher the popularity. For a better understanding, we will change the labels from the channel names to their view and subscriber counts.

We analyzed if the channels with the most collaborations that stand out the most have higher popularity indices. For example, we may find that

channels like PewDiePie[1], T-Series[2], or MrBeast[3], known for their high subscriber counts, also have significant numbers of collaborations. Conversely, we might see that some channels with numerous collaborations, such as smaller niche content creators, do not necessarily have high popularity despite their extensive network connections.

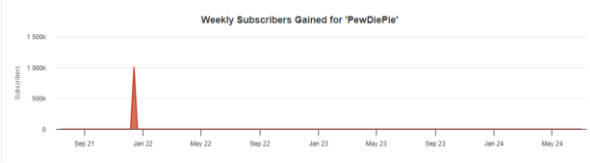

Figure 5. Weekly Subscribers Gained for “PewDiePie"

Figure 6. Weekly Subscribers Gained for “MrBeast”

Our hypothesis is that channels with more collaborations tend to be more popular, as collaborations can expose creators to wider audiences and different viewer demographics. However, there may be exceptions, such as highly popular channels that maintain their status through other means, like unique content or loyal fan bases, without relying heavily on collaborations.

By examining the correlation between collaboration frequency and popularity, we can understand the impact of networking on YouTube creators' success and provide insights into effective strategies for growing a YouTube channel.

So far, we have analyzed a network of YouTube creators connected by video collaborations. In this section, we will use the results of the analysis to find an answer to the following question:

Does the frequency of collaborations on YouTube correlate with a channel's popularity, and does it impact their overall success?

To answer this question, we will consider the view count and subscriber data of the channels involved. Popularity is defined as the number of views and subscribers a channel has. We will identify the nodes with the highest degrees (indicating numerous collaborations) and compare their view and subscriber counts to determine if more collaborations lead to greater popularity.

First, we will re-draw the network, focusing on the central component where there are the most nodes and the most connections. In the visual representation, we will use a heatmap-style color palette to represent the popularity of the channels: the darker the color, the higher the popularity. For a better understanding, we will change the labels from the channel names to their view and subscriber counts.

We will analyze if the channels with the most collaborations that stand out the most have higher popularity indices. For example, we may find that channels like PewDiePie, T-Series, or MrBeast, known for their high subscriber counts, also have significant numbers of collaborations. Conversely, we might see that some channels with numerous collaborations, such as smaller niche content creators, do not necessarily have high popularity despite their extensive network connections.

Our hypothesis is that channels with more collaborations tend to be more popular, as collaborations can expose creators to wider audiences and different viewer demographics. However, there may be exceptions, such as highly popular channels that maintain their status through other means, like unique content or loyal fan bases, without relying heavily on collaborations.

By examining the correlation between collaboration frequency and popularity, we can understand the impact of networking on YouTube creators' success and provide insights into effective strategies for growing a YouTube channel.

Collaborations on YouTube contribute significantly to the growth of channels' popularity, helping them achieve success. However, it's not the only factor in success, and some creators can achieve significant popularity through other strategies.

5. DISCUSSION

We have analysed a network of YouTube creators connected by video collaborations. Our analysis aimed to understand the impact of collaborations on the popularity of these creators. We found that, in many cases, frequent collaborations can enhance a creator's popularity, but it is not the sole factor determining success. Some creators achieve high levels of popularity with minimal collaborations, while others, despite numerous collaborations, struggle to reach the top. Overall, having a substantial number of high-quality collaborations can provide a significant boost to a creator's popularity on YouTube.

6. REFERENCES

[1] [n.d.]. Social Network Analysis Courses. https://github.com/sna-unipi

[2] [n.d.]. Pandas Documentation. https://pandas.pydata.org/docs/

[3] [n.d.]. NumPy Documentation. https://numpy.org/doc/stable/reference/index.html.

[4] [n.d.]. YouTube Data API. https://developers.google.com/youtube/v3

[5] [n.d.]. Software for Complex Networks - NetworkX 3.3 documentation. https://networkx.org/documentation/stable/

[1] Stats: PewDiePie - https://socialblade.com/youtube/channel/UC-lHJZR3Gqxm24_Vd_AJ5Yw/monthly

[2] T-Series - https://socialblade.com/youtube/channel/UCq-Fj5jknLsUf-MWSy4_brA/monthly

[3] MrBeast - https://socialblade.com/youtube/user/mrbeast6000/monthly