The YouTube Collaboration Network: Structural Patterns and Popularity Dynamics

Bhavya Jain (12240420)
Bhoomi Goyal (12240430)
Kriti Arora (12240880)
Soni Singh (12241780)

March 28, 2025

Introduction

This project addresses the problem of understanding how content creators on YouTube collaborate across genres, languages, and networks, and how these collaborations influence popularity, and engagement. We implement network analysis techniques on a youtube collaboration dataset, which represents collaborations between YouTubers as a directed, weighted graph. The need for this project arises from the lack of accessible, large-scale analysis of content-based collaboration patterns and their impact on visibility and community structure. By conducting this study, we aim to provide insights into genre-based clustering, language effects, network centrality, and influencer identification in online media ecosystems.

Objectives

The main objectives of this project are:

- **Collaboration Patterns:** Analyze how genre and language influence collaborations using graph-based techniques.
- **Collaboration Benefits:** Measure view count changes pre- and post-collaboration.
- **Network Properties:** Examine the largest connected component, node degree distribution, and collaboration density.
- **High-Benefit Creators:** Identify which creators gain the most from collaborations.
- **Community Detection:** Use modularity and Louvain algorithms to find clusters.
- **Influence Spread:** Apply diffusion models like:
  - *Independent Cascade Model (ICM)*
  - *Linear Threshold Model (LTM)*
  - *SIR & SIS Models*

Methodology

Dataset

The dataset has been taken from the results of the paper, **Collaborations on YouTube: From Unsupervised Detection to the Impact on Video and Channel Popularity, C. Koch, M. Lode, D. Stohr, A.** in which 2.4 years of video content and 3 months of youtube channel stats have been scrapped. The collaborations are detected using CATANA, a face recognition system on the basis of co-appearance in videos.

Nodes	Edges

{
"id": 0,
"label": "UCm1LjO7mpzb68Q71PjGleWQ",
"cluster": "1621",
"network": "None"
}
	{
	"source": 3619,
	"target": 464,
	"cluster": "5601",
	"weight": 1,
	"videos": "lAgzbO1QRtA"
	}

More file details are as follows: The direction here is from guest youtuber to host youtuber. The number of collaborations is used as edge weights.

- Tracks channel metrics (views, subscribers, comments, collabora- tions).
- Language data for channels.
- Direct creator collaborations.
- Video-level views and collaborations.
- Video and channel mentions.
- Graph of collaborations.

The project will be carried out in the following steps:

YouTube Collaboration Network Analysis

Methodology:

- Compute centrality measures:
  - **Degree Centrality:** Measures collaboration frequency.
  - **Betweenness Centrality:** Identifies bridge channels connecting communities.
  - **PageRank:** Determines influential collaborators.
  - **Geodesic Distance:** Use geodesic distance to find shortest collaboration paths.
- Detect clusters using Louvain or Girvan-Newman algorithms. Since we are having a ground truth structure. We can implement different definitions of community detection and find out the most natural groups in which they can be categorized. We will also see a local spectral clustering method to enhance community detection.
- Analyze collaboration weights to assess popularity impact.

Collaboration Impact on Popularity & Growth

Methodology:

- Compare pre- and post-collaboration growth using **viewCount_gradient** and **subscriberCount_gradient**
- Compare content categories (Entertainment, Gaming, Music, Education) for post-collab growth.
- Simulate network fragmentation by removing high-centrality nodes.

Content-Based Collaboration Patterns

Definition: Understanding how YouTubers collaborate based on content category and language preferences.
Methodology:

- Count collaborations per category and detect isolated genres.
- Identify cross-category collaborations (e.g., Gaming + Music).
- Analyze language-based collaboration trends.
- Detect multilingual creators collaborating across languages.

Most Frequent & Unusual Collaborations

Definition: Identifying common and unexpected collaboration patterns to analyze mutual benefits.
Methodology:

- Rank frequent collaboration pairs.

- Assess whether both creators benefit equally in terms of growth.
- Are there YouTubers who appear in many collaborations but have low actual engagement means views?

Expected Outcomes

- **Key Influencers and Collaborators:** Identify central creators, influential bridges, and collaboration hubs.
- **Language and Genre Dynamics:** Analyze whether language influences genre preferences and collaboration frequency.
- **Network Modeling:** Develop structured, weighted graphs to study collaboration patterns.
- **Content Marketing Insights:** Provide strategies based on genre and language collaboration trends.
- **Collaboration Impact:** Measure how partnerships affect creator growth and audience engagement.
- **Influence Spread:** Study mention-based influence and indirect creator interactions.
- **Visualization of Trends:** Generate graphs to showcase multi-collaborations, key clusters, and influential creators.
- **Quantitative Analysis:** Use filtered metrics for clear insights into the collaboration network.

Future work

- **Dynamics:** Our graph is not dynamic in any sense. We will make some more graphs in different timelines and scrape the statistics data for the timeline and study the growth.
- **Language and Genre Dynamics:** Analyze whether language influences genre preferences and collaboration frequency.
- **Geographical Metadata:** We can add geographical metadata to further analyze trends based on nationality and regional interests and how political opinions influence popularity.

References

- Collaborations on YouTube: From Unsupervised Detection to the Impact on Video and Channel Popularity, C. Koch, M. Lode, D. Stohr, A.
- Defining and Evaluating Network Communities based on Ground-truth
- Youtube Universe
- Assessing collaboration networks in educational research: A co-authorship-based social network analysis approach
- The Evolution of Trade and Scientific Collaboration Networks in the Global Wine Sector: A Longitudinal Study Using Network Analysis
- Scientific collaboration networks.