HUTECH UNIVERSITY

FACULTY of INFORMATION TECHNOLOGY

Unveiling Collaboration Patterns in V-pop Through YouTube Music: A Strategic Graph-Based Framework for Insights, Community Detection, & Link Predictions

Final Project in Social Network Analysis Course

Module Code	**CMP1048**
Instructor	**Nhat-Tung Le**
Student	**Hoang-Khang Nguyen**
Student ID	**2186400244**

Ho Chi Minh City, January 7, 2025

HUTECH UNIVERSITY
FACULTY of INFOMATION TECHNOLOGY

o0o

Final Project in Social Network Analysis Course

Unveiling Collaboration Patterns in V-pop Through YouTube Music: A Strategic Graph-Based Framework for Insights, Community Detection, & Link Predictions

Instructor: **Nhat-Tung Le**

Student: **Hoang-Khang Nguyen**

Student ID: **2186400244**

Ho Chi Minh City, January 7, 2025

Contents

List of Figures

List of Tables

Acronyms & Abbreviations

Table 1: List of Abbreviations and their Definitions

Abbreviation	**Definition**
V-pop	Vietnamese Pop Music
SNA	Social Network Analysis
LPA	Label Propagation Algorithm
Louvain	Louvain Modularity Optimization Algorithm
GCN	Graph Convolutional Network
PR	PageRank
AA	Adamic-Adar Index
CN	Common Neighbors
PA	Preferential Attachment
Jaccard	Jaccard Similarity Index
RA	Resource Allocation Index
SimRank	Structural Similarity Rank
Modularity	Measure of Community Structure Strength

Chapter 1 OVERVIEW

THE proliferation of social media platforms has revolutionized how artists connect with their audiences and collaborate with one another. In particular, YouTube [Sir+23] and its music-focused counterpart, YouTube Music, have emerged as dominant platforms for mainstream V-pop artists. With its vast library of music content and unparalleled reach, YouTube Music stands out compared to other streaming platforms like Spotify and Apple Music. Unlike its competitors, YouTube Music provides not only studio-produced tracks but also live performances, remixes, and fan-made content, creating an ecosystem that fosters engagement and discovery[JST24].

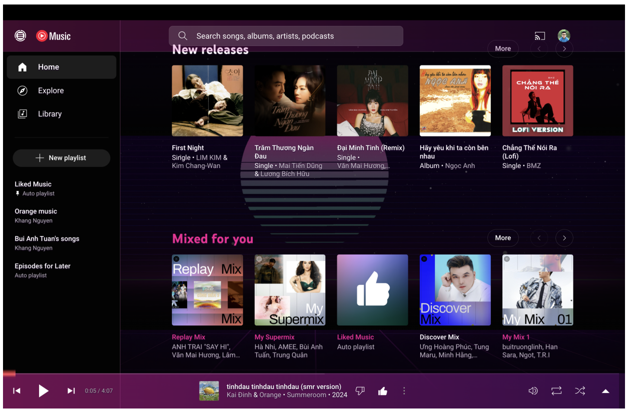

Figure 1.1: Youtube Music application

IN this project, we analyze the network of V-pop artists, modeled as a directed graph where nodes represent the official YouTube channels of the artists, and edges capture the collaborative relationships between them. The edge weights denote the number of directed collaborations in official music projects, making this graph both directed and weighted. Such a representation allows us to explore the intricate dynamics of the V-pop ecosystem and unveil insights into collaboration patterns and artist influence.

1.1 Objectives

1.1.1 Network Metrics and Visualization

TO understand the role and influence of each artist within the network, we computed key centrality metrics, including:

- **Weighted-degree centrality**: Quantifies an artist's overall collaborative impact.
- **Closeness centrality**: Measures how quickly an artist can connect to others in the network.
- **Betweenness centrality**: Highlights artists who act as bridges in the collaboration network.
- **Clustering centrality**: Explores the local density of collaborations around each artist.
- **PageRank**: Assesses an artist's importance based on the structure of incoming collaborations.

Additionally, we visualized the degree distribution and edge weight distribution, providing a comprehensive view of the network's structural properties and identifying outliers that signify exceptional artists or unique collaboration patterns.

1.1.2 Community Detection

TO reveal collaboration subgroups within the V-pop ecosystem, we applied three state-of-the-art community detection algorithms:

- **Louvain Method**: Known for its computational efficiency and its focus on maximizing modularity.
- **Girvan-Newman Algorithm**: A hierarchical approach based on edge betweenness centrality.
- **Label Propagation Algorithm (LPA)**: A fast, semi-randomized method that labels nodes iteratively.

The detected communities were compared in terms of modularity to evaluate their effectiveness. These insights provide a lens through which to understand the formation of collaboration cliques and their impact on the broader V-pop network.

1.1.3 Link Prediction

A KEY focus of our project is to forecast potential future collaborations among V-pop artists. Using metrics such as:

- **Adamic-Adar Index**, which emphasizes shared neighbors with lower degree nodes.
- **SimRank**, a recursive measure of similarity based on structural equivalence.
- **Preferential Attachment**, which predicts connections based on node degrees.
- **Jaccard Similarity**, which measures the overlap of neighborhoods between nodes.

These methods not only provide predictive insights into upcoming collaborations but also uncover latent patterns in how artists connect within the industry.

1.2 Relevance to the V-pop Ecosystem

THE V-pop industry, particularly its presence on YouTube and YouTube Music, has witnessed exponential growth in the last decade. Unlike traditional streaming platforms, YouTube enables V-pop artists to reach a global audience without intermediary barriers, fostering international collaborations and viral phenomena. This dynamic has been pivotal in shaping the identity of modern V-pop, with artists like Sơn Tùng M-TP, Jack (J97), and Hoàng Thùy Linh leveraging the platform to achieve mainstream success.

The ability to analyze and model such a network provides critical insights into:

- **Trends in collaboration**: Helping producers and managers identify potential partnerships.
- **Influence dynamics**: Which could be used to target marketing campaigns more effectively.
- **Emergent communities**: Offering a data-driven foundation for genre-based or regional segmentation within V-pop.

By linking our findings to the overarching trends in YouTube Music, we aim to contribute to a deeper understanding of the platform's role in the globalization of V-pop. YouTube's emphasis on music video engagement, combined with its integration of fan-generated content, positions it as a unique platform for artistic growth compared to Spotify's playlist-driven approach or Apple Music's premium-only model.

✍ **Note.**

IN summary, this project not only highlights the collaborative structure of V-pop artists on YouTube but also serves as a bridge for future studies into music industry analytics. The methodologies employed here have the potential to be adapted for other genres and platforms, further enriching the study of music collaboration networks.

■

Chapter 2 Crawl Data, Preprocessing, and Network Construction

THE construction of a meaningful social network of V-pop artists requires a systematic process of data crawling, preprocessing, and graph-based modeling. This chapter outlines the methodological steps used to extract data from YouTube, preprocess it, and construct the directed weighted network [Pro24].

2.1 Identifying Mainstream V-pop Artists

To focus on the most impactful and active artists, we compiled a list of mainstream V-pop artists using the following methodologies:

a. **Seed Artists:** Using an initial set of well-known artists, sourced from trending charts and playlists on YouTube Music.

b. **Expansion:** Expanding the list iteratively by analyzing related artists from charts, similar artist suggestions, and collaborations.

c. **Search Queries:** Conducting targeted searches (e.g., *"top Vietnamese pop artists," "trending V-pop," "rap viĬec7t"*).

2.1.1 Artist Identification Pipeline

The artist identification process involved a multi-step pipeline:

This process resulted in a comprehensive list of over 250 unique V-pop artists.

2.2 Identifying Official YouTube Channels

Identifying official YouTube channels was crucial to ensure data accuracy. Verified or official artist channels were prioritized. The identification process followed these steps:

- **Channel Search:** Using YouTube's API and `yt-dlp` to search for official channels based on artist names.
- **Verification:** Validating channels through metadata, including "verified" badges and "official artist channel" tags.
- **Filtering:** Excluding unofficial channels or fan pages by analyzing channel descriptions and upload patterns.

2.3 Tools Used: YouTubeDL and YTMusic

Two key tools were extensively used for crawling and analyzing YouTube data:

2.3.1 YouTubeDL

`YouTubeDL` is a command-line tool that allows efficient extraction of video metadata and content from YouTube [Dev25b]. Its key features include:

- Fetching detailed metadata for videos, channels, and playlists.
- Supporting search queries to locate artist channels.

Table 2.1: Steps for Identifying Mainstream V-pop Artists

Step	Description
Initial List	Seed list derived from well-known artists, curated manually and verified for relevance.
YouTube Music Charts	Extracted artist names from trending songs, videos, and artist charts in Vietnam.
Search Queries	Used YouTube Music API to search for artists across multiple genres and categories, filtering for relevance.
Similar Artists	Iteratively added artists by exploring "similar artists" suggestions from YouTube Music.
Validation	Cleaned and validated the list to ensure all identified names correspond to V-pop artists.

Table 2.2: Key Steps for Identifying Official YouTube Channels

Step	Description
Channel Search	Query YouTube using artist names combined with keywords like "official" or "official channel."
Verification	Check for verified badges or metadata indicating the channel type.
Filtering	Analyze descriptions and ensure the uploaded content aligns with the artist's identity.

- Allowing flexible filtering to identify specific types of videos.

Compared to the YouTube API, `YouTubeDL` offers:

- No strict API key or quota limits.
- Simpler setup for quick data extraction.
- Robust handling of dynamic changes in YouTube's website structure.

2.3.2 YTMusicAPI

`YTMusicAPI` is a Python library for interacting with YouTube Music. It was specifically used to analyze music charts and playlists [Dev25a]. Key advantages include:

- Accessing YouTube Music-specific metadata, such as charts, artist details, and playlists.
- Filtering results based on regions (e.g., Vietnam) to focus on V-pop content.
- Supporting operations like retrieving similar artists and related tracks.

Compared to the YouTube API, `YTMusicAPI` provides more tailored functionality for music-related data and requires minimal authentication setup.

2.4 Extracting Official Music Products

Official music products were identified by filtering uploaded videos. This process focused on classifying videos using specific patterns in titles and descriptions.

2.4.1 Key Patterns and Filters

- **Inclusion Keywords:** *"official MV," "official music video," "lyric video," "visualizer," "audio."*
- **Exclusion Keywords:** *"behind the scenes," "reaction," "teaser," "trailer," "karaoke."*

Table 2.3: Classification Rules for Music Products

Rule Type	Examples of Keywords
Inclusion	*official music video, MV, lyric video*
Exclusion	*reaction, teaser, trailer, behind the scenes*

2.5 Identifying Collaborations

Artist collaborations were identified by analyzing video titles and descriptions. Collaboration detection involved:

- **Patterns in Titles:** Using regular expressions to detect keywords such as "ft.," "feat.," "x," "featuring."
- **Credit Analysis:** Extracting names from structured credits or descriptions, such as "performed by" or "featuring."
- **Cross-checking:** Validating extracted names against the curated dataset of V-pop artists.

2.6 Building the Network

The collaborative network was modeled as a directed weighted graph where:

- **Nodes:** Represent artists.
- **Edges:** Represent collaborations, with weights indicating the frequency of collaboration.

Network construction involved iterating over detected collaborations and updating edge weights to reflect the number of collaborations between artists. The resulting graph was saved in `GEXF` format for further analysis and visualization.

2.7 Saving and Visualizing the Network

The network was saved in `GEXF` format, preserving node and edge attributes for further analysis. Visualization was performed using `networkx` and `matplotlib` to highlight key nodes and edges, emphasizing central figures in the V-pop ecosystem.

Table 2.4: Regular Expression Patterns for Collaboration Detection

Pattern	**Example**
ft., feat., featuring	*Artist A ft. Artist B, Artist A featuring Artist B*
x, X	*Artist A x Artist B*
Performed by	*Performed by Artist B*

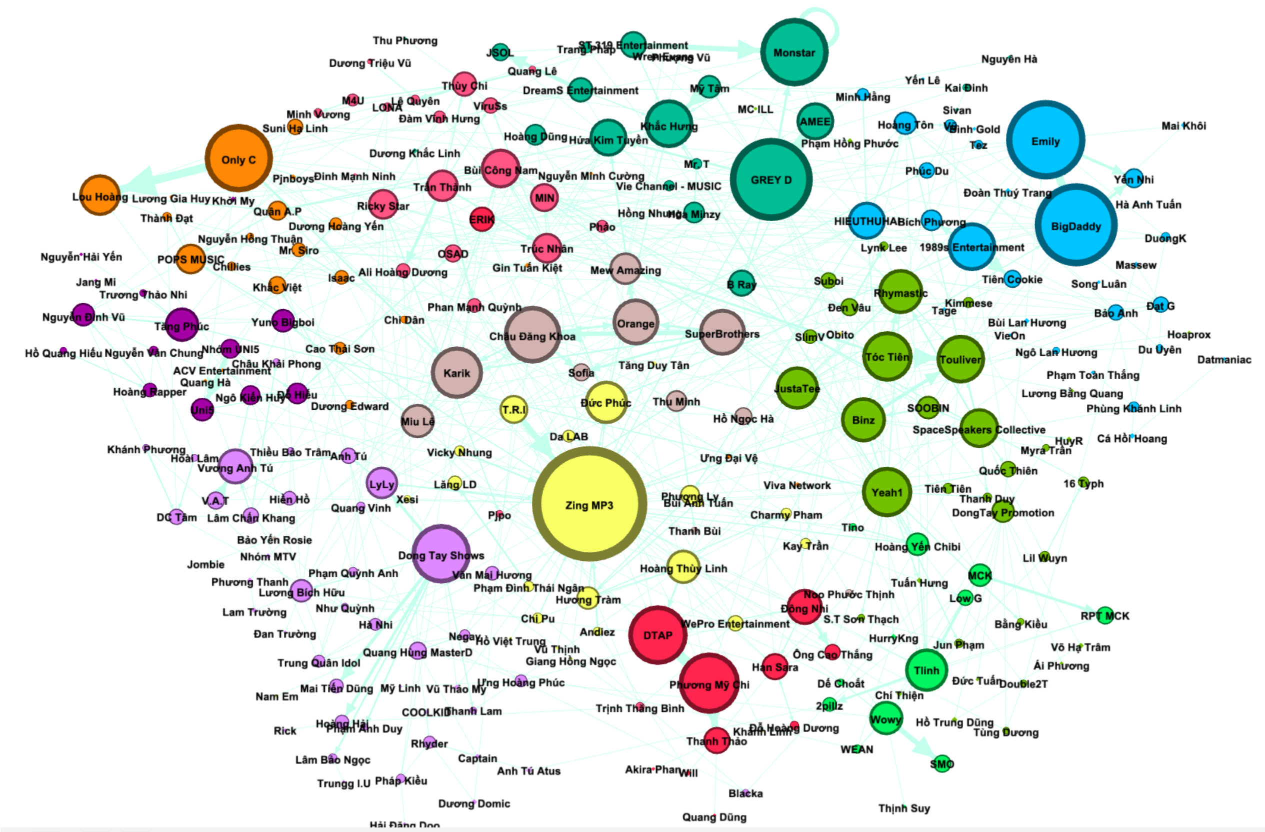

Figure 2.1: Network Construction

Table 2.5: Network Construction Overview

Step	**Description**
Node Creation	Each unique artist is added as a node in the graph.
Edge Addition	Directed edges are added between artists based on collaborations.
Weight Assignment	Edge weights represent the number of collaborations.

Chapter 3 Theory and Methodology

This chapter delves into the theoretical foundation and methodological framework underpinning the analysis of the V-pop collaborative network. By combining mathematical rigor with practical considerations, the following sections provide a comprehensive overview of graph metrics and their applications to the dataset.

3.1 Social Network Metrics

3.1.1 Graph Density

Graph density measures how "complete" a graph is, representing the ratio of the actual number of edges to the maximum possible number of edges.

Undirected Graphs

For undirected graphs, density is calculated as:

$$D = \frac{2|\mathcal{E}|}{|\mathcal{V}|(|\mathcal{V}| - 1)}$$

where:

- $|\mathcal{E}|$: The number of edges in the graph.
- $|\mathcal{V}|$: The number of vertices (nodes) in the graph.

Directed Graphs

For directed graphs, the formula is:

$$D = \frac{|\mathcal{E}|}{|\mathcal{V}|(|\mathcal{V}| - 1)}$$

Here, there is no factor of 2 because edges are directed, and each pair of nodes can have up to two edges (one in each direction).

✍ **Note.** Graph density is particularly useful for understanding the overall connectivity of the V-pop artist network. A higher density suggests a closely-knit community, which may indicate frequent collaborations between artists [Tùn25]. ■

3.1.2 Degree Centrality

Degree centrality quantifies the connectivity of a node in the network, indicating how many direct connections (edges) it has.

Undirected Graphs

For undirected graphs:

$$C_D(v) = \frac{\deg(v)}{|\mathcal{V}| - 1}$$

where $\deg(v)$ is the degree of node v, representing the number of edges connected to it.

Directed Graphs

For directed graphs:

$$C_D^{\text{in}}(v) = \frac{\deg^{\text{in}}(v)}{|\mathcal{V}| - 1}, \quad C_D^{\text{out}}(v) = \frac{\deg^{\text{out}}(v)}{|\mathcal{V}| - 1}$$

where:

- $\deg^{\text{in}}(v)$: The in-degree of node v (number of incoming edges).
- $\deg^{\text{out}}(v)$: The out-degree of node v (number of outgoing edges).

✍ **Note.** In the context of V-pop artists, degree centrality helps identify artists with the highest number of direct collaborations, reflecting their popularity or central role in the network. ■

3.1.3 Closeness Centrality

Closeness centrality measures how quickly information spreads from a node to all other nodes in the network.

$$C_C(v) = \frac{|\mathcal{V}| - 1}{\sum_{u \in \mathcal{V}, u \neq v} d(v, u)}$$

where:

- $d(v, u)$: The shortest path distance between nodes v and u.
- $|\mathcal{V}|$: The total number of nodes in the graph.

✍ **Note.** Artists with high closeness centrality are well-positioned to disseminate information or influence others quickly within the V-pop network. ■

3.1.4 Betweenness Centrality

Betweenness centrality quantifies the importance of a node as a bridge for information flow between other nodes.

$$C_B(v) = \sum_{s \neq v \neq t \in \mathcal{V}} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where:

- σ_{st}: The total number of shortest paths between nodes s and t.
- $\sigma_{st}(v)$: The number of those shortest paths that pass through node v.

✍ **Note.** High betweenness centrality indicates that an artist acts as a critical intermediary in the network, potentially influencing collaborations or the spread of trends. ■

3.1.5 Clustering Coefficient

The clustering coefficient measures the tendency of a node's neighbors to form a triangle (a fully connected subgraph).

Undirected Graphs

For undirected graphs:

$$C(v) = \frac{2T(v)}{\deg(v)(\deg(v) - 1)}$$

where $T(v)$ is the number of triangles involving node v.

Directed Graphs

For directed graphs:

$$C(v) = \frac{T(v)}{\deg^{\text{in}}(v)\deg^{\text{out}}(v) - \deg^{\text{loop}}(v)}$$

where:

- $\deg^{\text{loop}}(v)$: The number of self-loops at node v.

✍ **Note.** A high clustering coefficient suggests that an artist's collaborators are likely to collaborate with each other, forming tightly-knit communities within the V-pop network. ■

3.2 Community Detection Algorithms

3.2.1 Louvain Algorithm

The Louvain algorithm maximizes modularity in a hierarchical manner:

a. Initialize each node as its own community.
b. Repeat until no improvement:
   - For each node, move it to the community that maximizes modularity.
c. Aggregate nodes in the same community into a single node and repeat.

3.2.2 Girvan-Newman Algorithm

The Girvan-Newman algorithm detects communities by iteratively removing edges with the highest betweenness centrality:

a. Compute betweenness centrality for all edges.
b. Remove the edge with the highest betweenness centrality.
c. Repeat until the graph is divided into the desired number of communities.

✍ **Note.** Community detection algorithms provide insights into the structure of the V-pop artist network, revealing clusters of frequent collaborations. ■

3.2.3 Label Propagation Algorithm (LPA)

LPA is a community detection algorithm that assigns labels to nodes iteratively based on the labels of their neighbors. The process continues until label propagation stabilizes.

a. Initialize each node with a unique label.
b. For each node, update its label to the most frequent label among its neighbors.
c. Repeat until no label changes occur.

✍ **Note.** LPA is computationally efficient and works well for large-scale networks. It highlights overlapping communities where nodes can belong to multiple groups. ■

3.3 Similarity Metrics for Link Prediction

3.3.1 Common Neighbors

The common neighbors metric counts the number of shared neighbors between two nodes:

$$CN(u, v) = |N(u) \cap N(v)|$$

where $N(u)$ and $N(v)$ are the neighbor sets of nodes u and v.

✍ **Note.** This metric assumes that nodes with many common neighbors are more likely to be linked. ■

3.3.2 Adamic-Adar Index

The Adamic-Adar index weighs common neighbors inversely by their degree:

$$A(u, v) = \sum_{w \in N(u) \cap N(v)} \frac{1}{\log(\deg(w))}$$

where $\deg(w)$ is the degree of node w.

✍ **Note.** This metric emphasizes rare or exclusive connections by downweighting highly connected nodes. ■

3.3.3 Jaccard Similarity

Jaccard similarity measures the proportion of shared neighbors:

$$J(u, v) = \frac{|N(u) \cap N(v)|}{|N(u) \cup N(v)|}$$

3.3.4 Cosine Similarity

Cosine similarity calculates the cosine of the angle between the neighborhood vectors of two nodes:

$$C(u, v) = \frac{|N(u) \cap N(v)|}{\sqrt{|N(u)| \cdot |N(v)|}}$$

✍ **Note.** Cosine similarity is useful for capturing proportional similarity, especially when node degrees differ significantly. ■

3.3.5 Preferential Attachment

Preferential attachment predicts links based on node degrees:

$$PA(u, v) = \deg(u) \cdot \deg(v)$$

✍ **Note.** This metric assumes that highly connected nodes are more likely to form new links. ■

Chapter 4 Analysis and Evaluation of Experimental Results

THIS chapter provides an in-depth analysis and evaluation of the experimental results derived from the V-pop collaborative network. By leveraging advanced social network analysis metrics, community detection algorithms, and link prediction techniques, the analysis aims to uncover significant patterns, influential artists, and potential future collaborations within the network.

The network is modeled as a directed and weighted graph, where nodes represent Vietnamese V-pop artists, and edges denote collaborations weighted by their frequency. This structure allows for the computation of detailed metrics, such as degree centrality, betweenness centrality, closeness centrality, and PageRank, which are instrumental in identifying key artists who act as hubs, bridges, or influencers in the network.

Furthermore, community detection algorithms, including Louvain, Girvan-Newman, and Label Propagation, are applied to uncover modular structures and tightly-knit groups of artists. These communities reflect distinct collaboration trends and highlight the dynamics of the V-pop music industry.

Finally, link prediction metrics, such as Adamic-Adar, Jaccard similarity, and Preferential Attachment, are used to predict potential collaborations. This approach offers insights into how the network might evolve and identifies opportunities for future partnerships among artists.

The following sections detail the results of these analyses, supported by visualizations and tables to illustrate key findings and provide a comprehensive understanding of the V-pop collaborative network.

4.1 Network Metrics

This section evaluates fundamental properties of the V-pop collaborative network, offering insights into its structural characteristics and connectivity.

4.1.1 Basic Metrics

Key statistics for the network include:

✍ **Note.** The network's low density indicates a sparse structure, consistent with the specialized nature of collaborations in the V-pop industry. The clustering coefficient suggests limited clustering, reflecting a network where collaborations often span across different groups rather than being confined within tight cliques. The large number of components highlights the fragmented nature of the network, with many disconnected artists or isolated groups. ■

4.1.2 Degree Analysis

Degree metrics provide critical insights into the connectivity of the network.

✍ **Note.** The presence of nodes with very high degrees (hubs) indicates that some artists, such as prominent figures in the V-pop industry, act as central hubs. However, the median and mean degree suggest that most artists maintain a modest level of connectivity, aligning with the hierarchical structure of the music industry. ■

4.1.3 Scale-Free Analysis

The power-law analysis reveals whether the network exhibits scale-free properties:

- **Power-Law Exponent:** 1.77

Table 4.1: Basic Metrics of the V-pop Collaborative Network

Metric	**Value**
Number of Nodes	254
Number of Edges	683
Graph Density	0.0106
Average Clustering Coefficient	0.0713
Number of Components	146
Largest Strongly Connected Component Size	107
Average Shortest Path in Largest Component	4.1786
Diameter in Largest Component	11
Reciprocity	0.0996

Table 4.2: Degree Statistics of the Network

Metric	**Value**
Maximum Degree	44
Median Degree	4
Mean Degree	5.38
Standard Deviation of Degree	5.19

- **KS Statistic:** 0.227
- **p-value:** < 0.0001

✍ **Note.** The low p-value suggests the network does not strictly follow a scale-free distribution, indicating that while hubs exist, the distribution of connectivity does not adhere entirely to the power-law behavior typically observed in scale-free networks. This could reflect the diverse collaboration patterns unique to the V-pop industry. ■

4.1.4 Clustering Coefficient

The average clustering coefficient of 0.0713 indicates limited triadic closure, where artists' collaborators rarely form direct collaborations among themselves. This suggests a network with bridging ties rather than densely interconnected subgroups.

4.1.5 Reciprocity

A reciprocity value of 0.0996 highlights the directed nature of the network, where mutual collaborations are relatively rare. This aligns with the hierarchical collaboration patterns in the music industry, where prominent artists often collaborate with emerging talents, but the reverse is less frequent.

4.2 Community Detection

Community detection methods provide a detailed understanding of the collaborative structures within the V-pop network. By partitioning the graph into meaningful clusters, we identify groups of artists who frequently collaborate, thereby uncovering the underlying patterns of creative alliances.

4.2.1 Community Detection Algorithms

The V-pop collaborative network was analyzed using several community detection algorithms, including Louvain, Girvan-Newman, Label Propagation Algorithm (LPA), and K-Means on graph embeddings. The effectiveness of these methods was evaluated based on modularity and computational efficiency.

✍ **Note.** The Louvain algorithm outperformed others in terms of modularity and runtime, making it the most suitable choice for the V-pop network. Its ability to optimize modularity efficiently reflects the hierarchical and organic clustering structures present in the

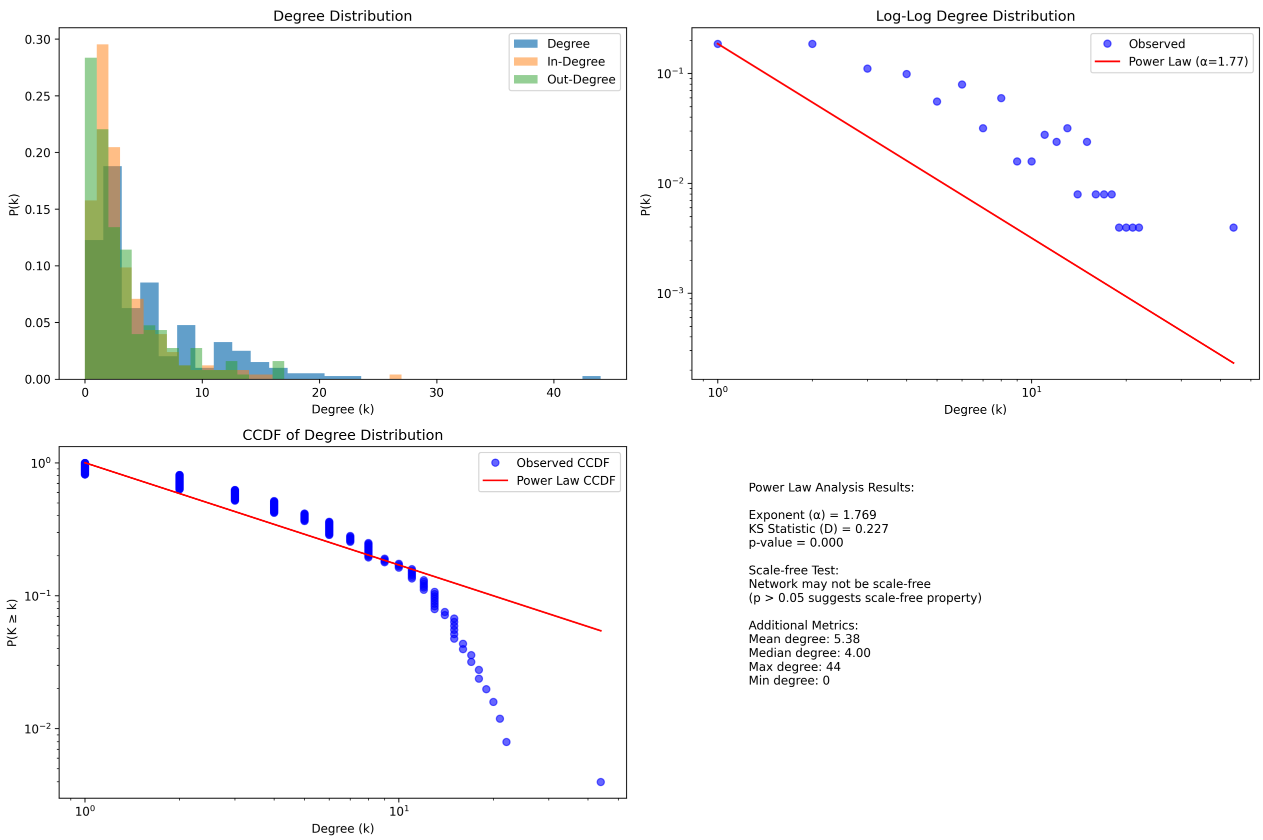

Figure 4.1: Degree distribution & Scale-free analysis

Table 4.3: Comparison of Community Detection Algorithms

Algorithm	**Number of Communities**	**Modularity**	**Runtime (s)**	**Key Characteristics**
Louvain	15	0.421	0.34	Hierarchical, fast convergence
Girvan-Newman	3	0.318	12.48	Computationally expensive, good for small graphs
LPA	25	0.372	0.22	Scalable, less modularity optimization
K-Means	20	0.358	0.75	Requires predefined clusters, embedding-dependent

network, aligning with the diverse but interconnected nature of V-pop collaborations. ■

4.2.2 Visualization of Louvain Partition

The Louvain algorithm's output is visualized in Figure 4.2, where communities are denoted by distinct colors. The size of each node represents its weighted degree, emphasizing key artists like Sơn Tùng M-TP, who act as central figures within their respective communities.

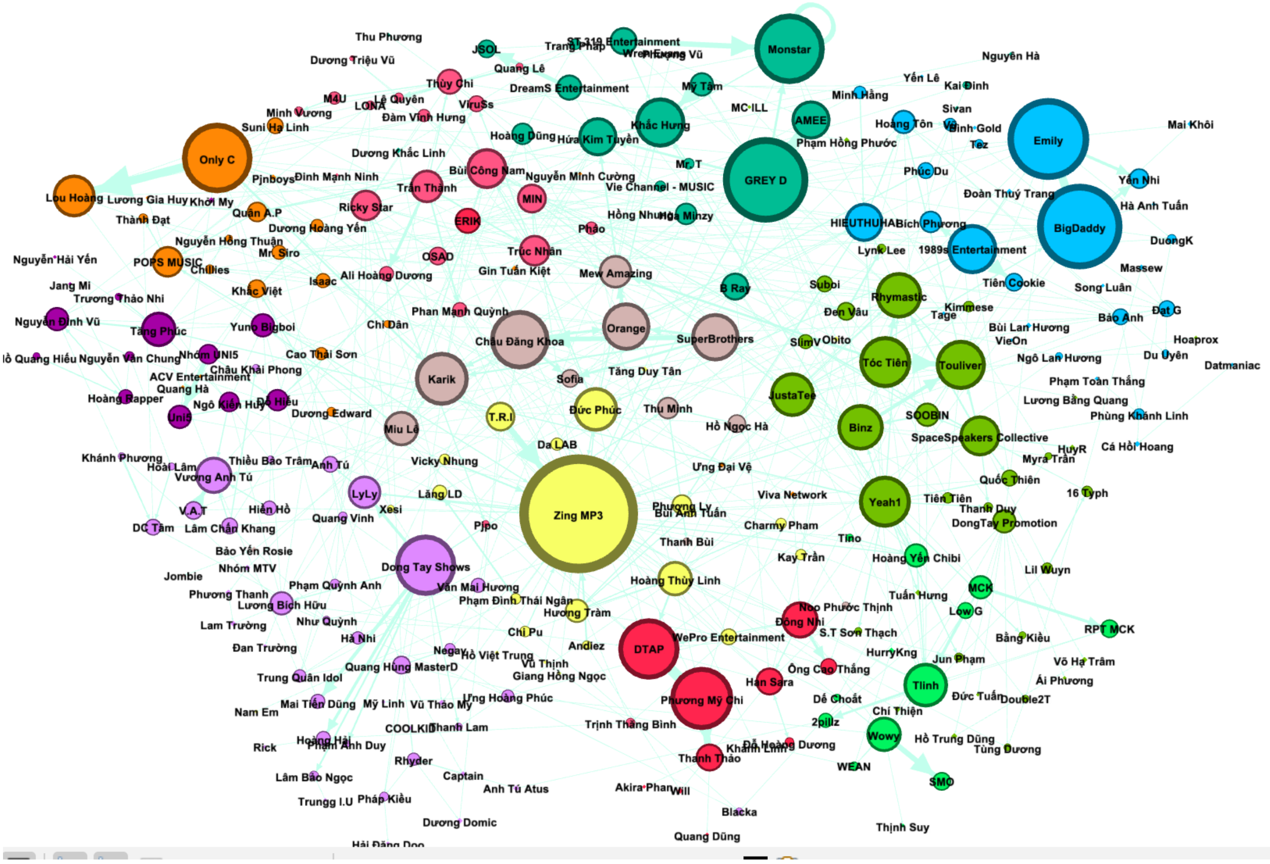

Figure 4.2: Community structure of the V-pop collaborative network detected using the Louvain algorithm.

4.2.3 Analysis of Community Structures

The community structures uncovered provide insights into collaborative tendencies in the V-pop industry:

- **Tightly-knit clusters:** These represent production teams or recurring partnerships, such as Zing MP3 collaborations with emerging artists.

- **Bridging communities:** Artists like Sơn Tùng M-TP and Hoàng Thùy Linh connect disparate clusters, reflecting their influence and cross-genre appeal.

✍ **Note.** The modular structure of the V-pop network highlights the fragmented yet interconnected nature of the industry. Louvain's success in identifying meaningful clusters underscores its utility for real-world social networks, particularly those exhibiting hierarchical and dynamic relationships. ■

4.3 Similarity Metrics and Link Prediction

Link prediction plays a crucial role in understanding potential future collaborations within the V-pop industry. By evaluating similarity metrics, we predict connections between artists and analyze their implications for the network's evolution.

4.3.1 Similarity Metrics

The following similarity metrics were applied to the V-pop network to assess the likelihood of future collaborations:

- **Jaccard Similarity:** Measures the overlap between two nodes' neighbors.

- **Adamic-Adar Index:** Weighs shared neighbors inversely by their degree.

- **Preferential Attachment:** Predicts links based on the product of node degrees.

- **Resource Allocation Index:** Similar to

Adamic-Adar but with a probabilistic interpretation.

- **Common Neighbors:** Counts the number of shared neighbors between two nodes.
- **Cosine Similarity:** Computes the cosine of the angle between neighborhood vectors.
- **SimRank:** Quantifies similarity based on the structural equivalence of nodes.

4.3.2 Performance Comparison

Table 4.4 summarizes the performance of each metric in predicting potential collaborations.

✍ **Note.** Preferential Attachment achieves the highest precision and recall, reflecting its ability to predict collaborations involving high-degree nodes. However, methods like Jaccard and SimRank show limited predictive power, indicating that shared neighbors alone are insufficient for capturing the dynamics of the V-pop network. ■

4.3.3 Practical Implications for V-pop

The results highlight distinct patterns in the network:

- **High-Degree Artists:** Preferential Attachment emphasizes prominent figures like Sơn Tùng M-TP and Hoàng Thùy Linh, who frequently collaborate with emerging artists.
- **Shared Neighborhoods:** Metrics like Common Neighbors and Adamic-Adar suggest potential collaborations within tightly-knit communities, such as those around Zing MP3.
- **Structural Equivalence:** SimRank's low performance indicates that V-pop collaborations are not purely based on structural equivalence but rather on other factors such as genre or popularity.

✍ **Note.** These insights provide actionable recommendations for stakeholders in the V-pop industry to foster strategic collaborations, leveraging high-degree hubs and tightly-knit communities to maximize creative output and market reach. ■

Table 4.4: Comparison of Link Prediction Methods

Method	**Precision**	**Recall**	**F1-Score**	**AUC**	**True Positives**
Jaccard	0.0	0.0	0.0	0.604	0
Adamic-Adar	0.022	0.022	0.022	0.610	3
Preferential Attachment	0.036	0.036	0.036	0.686	5
Resource Allocation	0.022	0.022	0.022	0.610	3
Common Neighbors	0.022	0.022	0.022	0.623	3
Cosine	0.022	0.022	0.022	0.620	3
SimRank	0.0	0.0	0.0	0.609	0

Chapter 5

Conclusion

THIS study provides a comprehensive analysis of the V-pop collaborative network using advanced graph-based methodologies. The results offer valuable insights into the structure and dynamics of the network, revealing key patterns of collaboration, community structures, and potential future partnerships. Below, we summarize the findings, discuss the limitations, and propose future directions.

5.1 Summary of Findings

The analysis highlights several significant insights into the V-pop network:

- The network is sparse, with a low density of connections, indicating that collaborations are selective and often span across distinct groups.
- Degree centrality and preferential attachment emphasize the prominence of key artists, such as Sơn Tùng M-TP and Hoàng Thùy Linh, who act as hubs.
- The Louvain algorithm effectively detects meaningful communities, reflecting production teams and recurring collaborations, and outperforms other community detection methods in terms of modularity and runtime.
- Link prediction metrics reveal the likelihood of future collaborations, with preferential attachment showing the strongest predictive power.

These findings provide actionable insights for stakeholders in the V-pop industry to optimize collaborations and foster growth.

5.2 Limitations and Challenges

Despite the successes, the study faced several challenges and limitations:

- **Data Availability:** The reliance on publicly available data limited the scope of analysis. Certain collaborations might not be reflected due to incomplete metadata or unverified sources.
- **Algorithm Scalability:** Methods like Girvan-Newman and SimRank proved computationally expensive, making them less suitable for large-scale networks.
- **Similarity Metrics:** Metrics such as Jaccard and SimRank underperformed in predicting links, highlighting the need for more nuanced models that account for genre, temporal trends, and artist preferences.
- **Subjectivity in Validation:** The lack of a ground truth for collaborations introduces subjectivity in validating the results of link prediction and community detection.

5.3 Future Directions

Building on the current findings, several future directions are proposed:

- **Incorporating Temporal Dynamics:** Analyzing the temporal evolution of the network can provide insights into how collaborations and communities change over time.
- **Genre-Based Analysis:** Grouping artists by genre could reveal genre-specific collaboration

patterns and aid in targeted marketing strategies.

- **Multilayer Networks:** Extending the analysis to multilayer networks, such as integrating social media and streaming data, could capture a more holistic view of the V-pop ecosystem.
- **Enhanced Link Prediction Models:** Incorporating machine learning models that use network embeddings and metadata could improve link prediction accuracy.
- **Expanding Dataset Coverage:** Crawling additional data sources, such as concert collaborations and producer credits, can enrich the network and reveal hidden patterns.

5.4 Implications for the Vietnamese Music Industry

This study highlights the following implications for the V-pop industry:

- **Strategic Collaborations:** Identifying key artists and communities enables producers and marketers to foster strategic collaborations that maximize impact.
- **Supporting Emerging Artists:** By analyzing connectivity patterns, stakeholders can identify and support emerging artists who demonstrate potential for growth within influential communities.
- **Cross-Genre Opportunities:** Insights into bridging nodes suggest opportunities for cross-genre collaborations, which can diversify the market and attract broader audiences.
- **Data-Driven Decision Making:** Leveraging network analysis tools allows for data-driven strategies in artist management, marketing campaigns, and content production.

✍ **Note.** This study underscores the potential of network analysis in understanding and shaping the V-pop industry. By addressing the identified limitations and pursuing future directions, this framework can be extended to other creative industries, fostering innovation and growth. ■

BIBLIOGRAPHY

[Sir+23] Hiba Siraj et al. "How Do YouTubers Collaborate? A Preliminary Analysis of YouTubers' Collaboration Networks." In: Mar. 2023.

[JST24] Loso Judijanto, Irwan Suryadi, and Rahman Tanjung. "Network Analysis of Research Collaboration in Skills-Based Leadership: A Bibliometric Approach." In: *West Science Business and Management* 2 (Mar. 2024), pp. 64–70. DOI: `10.58812/wsbm.v2i01.735`.

[Pro24] The Clever Programmer. *YouTube Data Collection and Analysis with Python.* `https://thecleverprogrammer.com/2024/06/03/youtube-data-collection-and-analysis-with-python/`. Accessed: January 8, 2025. 2024.

[Dev25a] th-ch Developers. *YouTube Music API: A Lightweight Unofficial YouTube Music Library.* `https://github.com/th-ch/youtube-music`. Accessed: January 8, 2025. 2025.

[Dev25b] youtube-dl Developers. *youtube-dl: A Command-Line Program to Download Videos from YouTube and Other Sites.* `https://github.com/ytdl-org/youtube-dl`. Accessed: January 8, 2025. 2025.

[Tùn25] Lê Nhật Tùng. *Social Network Analysis Project Repository.* `https://github.com/lenhattung/socialnetwork/tree/main`. Accessed: January 8, 2025. 2025.