

$$f : D \times D \times I \longrightarrow [0, 1]^{|\mathcal{R}_I|}$$

UNVEILING COLLABORATION PATTERNS
IN V-POP THROUGH YOUTUBE MUSIC:
A STRATEGIC GRAPH-BASED
FRAMEWORK FOR INSIGHTS, COMMUNITY
DETECTION, & LINK PREDICTIONS

Khang Nguyen-Hoang

HUTECH UNIVERSITY

January 7, 2025

INTRODUCTION

- V-pop has witnessed exponential growth on YouTube Music.
- Collaboration is a key driver of creative synergy and audience reach.
- This study employs graph-based methods to uncover collaboration patterns, community structures, and predict future connections.

network_visualization.png

- Degree, Closeness, Betweenness, and PageRank were calculated to identify key artists.
- Insights:
 - Sơn Tùng M-TP emerged as a central hub.
 - Hoàng Thùy Linh showcased high betweenness, indicating her bridging role.

Table 1: Top 5 Artists by Centrality Metrics

Artist	Degree	Closeness	Betweenness	PageRank
Sơn Tùng M-TP	44	0.89	0.15	0.12
Hoàng Thùy Linh	37	0.87	0.13	0.11
Grey D	35	0.85	0.11	0.10
DTAP	30	0.83	0.09	0.09
Big Daddy	28	0.81	0.08	0.08

COMMUNITY DETECTION

- Communities were detected using:
 - **Louvain Algorithm:** Most effective with a modularity score of 0.42.
 - Girvan-Newman and LPA performed relatively poorer in modularity and runtime.
- Louvain results align with production teams and recurring collaborations.

Table 2: Community Detection Comparison

Algorithm	Modularity	Runtime (s)
Louvain	0.42	2.3
Girvan-Newman	0.38	45.8
Label Propagation	0.35	1.9

- Metrics used for prediction:
 - Adamic-Adar: Focus on shared neighbors.
 - Preferential Attachment: Degree-based prediction.
 - Common Neighbors: Neighborhood overlap.
- Results reveal potential collaborations among emerging and established artists.

Table 3: Performance of Link Prediction Metrics

Metric	Precision	Recall	F1-Score
Adamic-Adar	0.62	0.58	0.60
Preferential Attachment	0.68	0.61	0.64
Common Neighbors	0.59	0.54	0.56

- Key insights:
 - Identification of hubs (e.g., Sơn Tùng M-TP, Hoàng Thùy Linh).
 - Community structures highlight collaboration trends.
- Limitations:
 - Reliance on publicly available data.
 - Scalability of computational methods.
- Future directions:
 - Temporal dynamics and genre-specific analysis.
 - Enhanced prediction models with network embeddings.