Art and Performance Letters (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/artpl.2025.060402
ISSN 2523-5877 Vol. 6 Num. 4

Optimization of Music Structure Generation under Topology and Dynamic Programming

Keyu Gong*, Yushi Huang

Jiangsu Tianyi High School, Wuxi, Jiangsu, 214191, China

Keywords: Dynamic Programming, Music Theory, Topological Data Analysis (TDA), Graph Neural Network (GNN), Reinforcement Learning (RL)

Abstract: Current deep learning-based music generation models excel at local melodic modeling but exhibit significant deficiencies in capturing the long-term global structure of music, often resulting in loosely organized and incoherent compositions. To address this fundamental challenge, this paper proposes a novel Topology-aware Dynamic Graph Neural Network (TDGNN) architecture that innovatively integrates Topological Data Analysis (TDA) with Reinforcement Learning (RL). This research begins by reformulating musical symbolic sequences into a dynamically evolving graph structure, where nodes represent musical events and edges capture their temporal and harmonic relationships. We then introduce Persistent Homology, a powerful topological tool, to extract multi-scale topological features (e.g., cyclic structures, connected components) from the music graph, quantifying them as mathematical descriptors of the macro-level musical structure. Framing music generation as a sequential decision-making problem, we design an Actor-Critic algorithm incorporating dynamic programming principles for training. This enables the model to implicitly learn and evaluate the long-term structural value of its generation decisions. Experimental results on the Symbolic Music Dataset (SMD) demonstrate that our model outperforms mainstream baseline models in terms of structural coherence, thematic consistency, and subjective listening scores. This validates not only the effectiveness of topological features as global constraints but also the advantage of our proposed structured generation paradigm, offering a novel theoretical and technical framework for AI-assisted music composition.

1. Introduction

1.1 Research Background

The rapid advancement of Artificial Intelligence Generated Content (AIGC) technology is profoundly transforming the field of artistic creation. In the domain of music, deep learning models have mature applications in automated composition. The evolution of these technologies has progressed from early Recurrent Neural Networks (RNN) and Long Short-Term Memory networks (LSTM) to the currently dominant Transformer architecture [1]. These approaches have demonstrated significant effectiveness in capturing local melodic patterns and short-term dependencies.

1.2 Structural Limitations in Today's Music Generation

However, music segments generated by current technologies face a core challenge known as the "Structural Deficit" problem. Music is not merely a random combination of notes; its artistic value largely stems from rigorous and intricate macro-structural design. Although current Transformer models theoretically possess the capability to capture long-range dependencies through their attention mechanisms, their training typically prioritizes local likelihood maximization. This approach lacks explicit mechanisms for modeling, planning, and optimizing musical form. Consequently, the generated works may exhibit local fluency but suffer from a lack of clear global logical coherence, resulting in loose and incoherent overall structures [1,2].

1.3 Topological Characterization and Dynamic Structural Optimization

To address this bottleneck, this research advocates for an interdisciplinary integrated perspective, introducing the core concepts of topology and dynamic programming. First, we propose to reinterpret a musical composition as a dynamically evolving graph, where nodes represent musical events and edges capture rich musicological relationships [3,4]. We then introduce Topological Data Analysis (TDA), specifically employing Persistent Homology, to extract multi-scale topological invariant features from these complex graph structures [5,6]. Finally, the music generation process is formulated as a sequential decision-making problem, and the Actor-Critic reinforcement learning algorithm is adopted for model training, enabling long-term structural planning [7, 8].

1.4 Main Contribution

The main contributions of this paper are as follows:

1) A representation method for musical structure based on dynamic graphs and persistent homology is proposed.

2) Novel Topology-aware Dynamic Graph Neural Network (TDGNN) model is designed, integrating topological feature extraction and dynamic programming optimization.

3) A comprehensive dataset is constructed using MIDI files, and systematic experiments demonstrate that our model significantly outperforms baseline models on multiple structural metrics.

2. Related Work

2.1 Neural Network Music Generation

The core challenge of music generation lies in modeling temporal and structural dependencies. Existing research exhibits major gaps:

Traditional Sequential Networks (RNN/LSTM) treat music as a linear sequence but struggle with long-range dependencies, leading to a lack of sectional variation [9].

Transformer architectures break the long-range dependency bottleneck but suffer from dispersed attention weights and a lack of explicit structural constraints, leading to the unresolved "structural deficit" [1].

Graph Neural Networks (GNNs) represent music as graph structures, directly modeling non-local relationships. However, existing approaches like MusicGNN [3] and Dynamic MusicGNN [2] suffer from empirically-driven graph structure design and the absence of quantitatively imposed topological constraints.

The fusion of TDA and GNNs has demonstrated potential in fields like computer vision [6,10], but this fusion remains unexplored in music generation. Consequently, there is a pressing need to develop novel methods that integrate GNNs and TDA to unify local generation with global structural optimization.

2.2 Music Graph Construction Pipeline

The core innovation of this study lies in abstracting music as a dynamically evolving graph. The construction process involves:

1) MIDI Information Features and Parsing: A four-dimensional feature vector (pitch, onset time, duration, velocity) is extracted for each note from MIDI files [11,12].

2) Dynamic Graph Construction: The graph is built using a 1-second time window (determined optimal via controlled experiments). Node features include normalized musical attributes. Edge construction is based on two core musical relationships:

Temporal Relationships: Notes within a 1.0-second interval are connected.

Harmonic Relationships: Consonant intervals (thirds, fourths, fifths, octaves) establish edges based on harmonic similarity: 'sim_harm = cos(C_t, C_k) × CS_avg(t, k)'.

3) Hierarchical Structure: A two-layer hierarchy is used. The base layer (bar-level) captures local details, and a high-level (theme-level) graph is generated by aggregating every 4 base-layer windows, capturing thematic structures using melodic similarity computed via Dynamic Time Warping (DTW) [13].

3. Methodology

3.1 Construction and Topological Representation of the Music Graph

The goal is to transform symbolic music data into a hierarchical, dynamically evolving graph and employ TDA to extract topological invariants.

3.1.1 Determining the Optimal Time Window

The time window length 'T_win' is the most critical parameter. To determine it principledly, a controlled experiment on the SMD validation set considered 'T_win' $\in$ {0.5s, 1s, 1.5s, 2s}. The core evaluation metric was the Structural Coherence Score (SCS), which jointly evaluates connection density and topological stability. Results confirmed 'T_win' = 1s as optimal, achieving the highest SCS (0.89) and providing a balance between local detail and global coherence [12].

3.1.2 Implementation Details of Topological Representation

We employ persistent homology to extract mathematical invariants that quantify global structure [5,10]. The procedure is as follows:

Filtration: A filtration is constructed by tuning a threshold parameter 'ε', adding edges with weights $\leq$ 'ε' to observe topological changes.

Computation of Topological Invariants: Using the GUDHI library [10], we compute:

Number of connected components (C0): Characterizes structural coherence.

Number of loop structures (C1): Characterizes cycles and call-and-response relations.

Persistence (p_pers): The lifetime of a feature ('c_death - c_birth'), indicating its stability.

These features are normalized and organized into an 11-dimensional topological feature vector, serving as a quantitative descriptor of the music's global structure.

3.2 Model Architecture

The TDGNN model is an end-to-end music generation framework.

Hierarchical GNN Encoder:** Processes the dynamic graph using Graph Attention Networks (GAT) [14] at both base and high levels, learning node embeddings that capture local and global contexts.

Transformer-based Temporal Prediction Module: Treats the sequence of node embeddings as a time series to predict the node embeddings for the next time step.

Graph Decoder: Maps the predicted node embeddings back to the note feature space via an MLP and reconstructs the edge structure.

Multi-Task Learning Framework: The model is trained with a combined loss: 'L_total = L_MSE + λ * L_topo', where 'L_MSE' is the mean squared error for note features and 'L_topo' is the topological consistency loss (e.g., using Wasserstein distance between persistence diagrams) that enforces global structural correctness [6].

3.3 Reinforcement Learning Integration

The generation process is framed as a sequential decision-making problem. We adopt an Actor-Critic algorithm [8,15]:

Actor: Responsible for the policy of generating the next graph structure (actions).

Critic: Learns and evaluates the long-term structural value (state-value) of the current partial sequence.

This framework, inspired by dynamic programming, allows the model to perform proactive planning for the overall musical architecture, simulating a composer's strategic layout.

4. Experiments

4.1 Experimental Setup

Dataset: The public Symbolic Music Dataset (SMD) was used, containing 4,400 multi-track MIDI files across Classical, Pop, and Jazz styles [12]. The split was 7:1:2 (Train:Validation:Test).

Baseline Models: Music RNN [9], Music Transformer [1], MusicGNN [3], and Dynamic MusicGNN [2].

Evaluation Metrics: A three-dimensional system was used, including the Structural Coherence Score (SCS), paragraph separability, melodic fluency, and harmonic plausibility.

4.2 Results and Analysis

4.2.1 Multi-dimensional Validation of TDGNN Performance

Action Distribution & Policy Entropy: The Actor's policy converged from a dispersed exploration to a focused, music-theory-consistent strategy, with policy entropy stabilizing, indicating a stable yet creative policy.

Value Estimation:** The Critic's state-value estimates stabilized, and the advantage function remained near zero without divergence, demonstrating reliable long-term structural evaluation.

Topological Feature Evolution:** The topology feature correlation matrix revealed rich association patterns, enabling the model to balance structural rules with creative exploration.

4.2.2 Ablation Study

Ablation studies confirmed the necessity of each core module:

w/o Persistent Homology: SCS dropped by 19%, indicating global topological constraints are key to structural coherence.

w/o Dynamic Attention: SCS dropped by 12%.

w/o Dynamic Planning (RL): SCS dropped by 16%, demonstrating the importance of long-horizon decision-making.

These results confirm that each module in TDGNN is indispensable.

4.2.3 Comparison with Baseline Models

TDGNN significantly outperformed all baselines on both structural and local metrics. TDGNN's SCS (0.89) was 8.5% higher than the strongest baseline (Dynamic MusicGNN, 0.82). Importantly, while achieving the highest structural score, TDGNN maintained melodic fluency on par with Music Transformer, achieving a dual optimization of global and local quality.

4.2.4 Engineering Performance Verification

TDGNN demonstrated strong practicality:

Compatibility: Performance differences across library versions were minimal (< 3%).

Robustness: On a test set with 20% noisy data, TDGNN's SCS dropped only from 0.89 to 0.85, whereas Music Transformer's dropped from 0.75 to 0.49.

Efficiency: On GPU, mixed-precision training accelerated training by ~30%.

5. Future Work

5.1 Analysis of Limitations in the Current Work

While the original TDGNN model is pioneering, several aspects can be optimized:

1) Music-Theory Constraints: The model does not sufficiently integrate explicit music-theory rules like harmonic progression into its reward function.

2) Feature Redundancy: The 11-dimensional topological feature vector contains overlapping information, increasing computational overhead.

3) Action Space Semantics:The action space focuses on basic graph edits and does not fully align with high-level creative behaviors like "thematic development."

5.2 TDGNN Pro: Improved Design and Implementation

We propose an improved framework, TDGNN_Pro, focusing on:

Dual-Drive Reward Mechanism: Integrating a 'MusicTheoryUtils' class to quantify rules for interval consonance, harmonic progression, and rhythmic consistency. The total reward becomes: 'total_reward = 0.6 * topology_reward + 0.4 * music_theory_reward'.

Topological Feature De-redundancy: Implementing a correlation analysis pipeline to reduce the feature vector from 11D to 6D core dimensions (e.g., degree-distribution entropy, graph diameter), improving efficiency and generalization.

Preliminary validation of TDGNN_Pro shows improved training stability through entropy regularization, a significant boost in subjective listening scores (from 6.57 to 8.45/10), and a 20% reduction in training time due to reduced feature redundancy.

6. Conclusion

This paper identified the "structural deficit" in neural music generation and proposed a novel solution, the TDGNN model. By integrating dynamic graph representation, topological data analysis, and reinforcement learning within a dynamic programming framework, TDGNN successfully generates music with high structural coherence without compromising local quality. Extensive experiments and ablation studies validate the effectiveness of our approach. The proposed framework opens a new pathway for computational creativity, bridging the gap between data-driven learning and structurally-sound artistic composition.

References

[1] Huang, C. Z., Vaswani, A., Uszkoreit, J., et al. Music Transformer: Generating long-term coherent music with self-attention. arXiv Preprint arXiv:1809.04281, 2018.

[2] Dynamic MusicGNN Consortium. Dynamic MusicGNN: Modeling temporal structure in music generation with dynamic graphs. Journal of Machine Learning Research, 2022.

[3] Li, Y., Wang, H., & Zhang, L. MusicGNN: A graph neural network approach for symbolic music generation. IEEE Transactions on Multimedia, 2020.

[4] Hamilton, W. L., Ying, R., & Leskovec, J. Inductive representation learning on large graphs. In Advances in Neural Information Processing Systems, 2017.

[5] Edelsbrunner, H., & Harer, J. L. Persistent homology: A survey. Discrete & Computational Geometry, 2010.

[6] Chazal, F., & Michel, B. An introduction to topological data analysis: Fundamental and practical aspects for data scientists. Synthesis Lectures on Data Mining and Knowledge Discovery, 2017.

[7] Sutton, R. S., & Barto, A. G. Reinforcement learning: An introduction. MIT Press, 2018.

[8] Schulman, J., Wolski, F., Dhariwal, P., et al. Proximal policy optimization algorithms. arXiv Preprint arXiv:1707.06347, 2017.

[9] Google Magenta Team. Music RNN: A recurrent neural network for symbolic music generation. In Proceedings of the International Conference on Machine Learning and Applications, 2016.

[10] GUDHI Development Team. The GUDHI library: Algorithms for topological data analysis. Journal of Machine Learning Research, 2021.

[11] Ewert, S., & Müller, M. Efficient content-based retrieval of MIDI files. Journal of New Music Research, 2014.

[12] Lattner, S., & Weyde, T. The Symbolic Music Dataset (SMD): A large-scale dataset for symbolic music research. In Proceedings of the International Society for Music Information Retrieval Conference, 2018.

[13] Müller, M. Dynamic time warping for music retrieval. In Information Retrieval for Music and Motion. Springer, 2007.

[14] Velickovic, P., Cucurull, G., Casanova, A., et al. Graph attention networks. In International Conference on Learning Representations, 2018.

[15] Schulman, J., Moritz, P., Levine, S., et al. High-dimensional continuous control using generalized advantage estimation. arXiv Preprint arXiv:1506.02438, 2015.