Synthesizing Composite Hierarchical Structure from Polyphonic Music Corpora

(/pdf?id=4vD5vibp7S)

Ilana Shapiro (/profile?id=~Ilana_Shapiro1), Ruanqianqian (Lisa) Huang (/profile?email=r6huang%40ucsd.edu), Zachary Novack (/profile?id=~Zachary_Novack1), Cheng-i Wang (/profile?email=cheng-i%40audioshake.ai), Hao-Wen Dong (/profile?id=~Hao-Wen_Dong1), Taylor Berg-Kirkpatrick (/profile?id=~Taylor_Berg-Kirkpatrick1), Shlomo Dubnov (/profile?id=~Shlomo_Dubnov1), Sorin Lerner (/profile?id=~Sorin_Lerner1)

07 Aug 2024 (modified: 09 Dec 2024) Submitted to AAAI 2025
Conference, Area Chairs, Senior Program Committee, Program Committee, Authors
Revisions (/revisions?id=4vD5vibp7S) BibTeX

Primary Keyword: Application Domains (APP) -> APP: Humanities & Computational Social Science
Secondary Keywords: Constraint Satisfaction and Optimization (CSO) -> CSO: Applications
TL;DR: We introduce a meta-representation of music structure, formalize structural summarization of a music corpus as a bi-level NP-hard combinatorial optimization problem, and solve with nested simulated annealing and the formal guarantees of SMT solvers.

Abstract:
Western music is an innately hierarchical system of interacting levels of structure, from fine-grained melody to high-level form. In order to analyze music compositions holistically and at multiple granularities, we propose a unified, hierarchical meta-representation of musical structure called the *structural temporal graph* (STG). For a single piece, the STG is a data structure that defines a hierarchy of progressively finer structural musical features and the temporal relationships between them. We use the STG to enable a novel approach for deriving a representative structural summary of a music corpus, which we formalize as a dually NP-hard combinatorial optimization problem. Our approach first applies simulated annealing to develop a measure of *structural distance* between two music pieces rooted in graph isomorphism. Our approach then combines the formal guarantees of SMT solvers with nested simulated annealing over structural distances to produce a structurally sound, representative *centroid* STG for an entire corpus of STGs obtained from individual pieces. To evaluate our approach, we conduct experiments showing that structural distance accurately differentiates between music pieces, and that our computed centroids encapsulate the overarching structure of a music corpus.

Supplementary Material: zip (/attachment?id=4vD5vibp7S&name=supplementary_material)
iThenticate Agreement: Yes, I agree to iThenticate's EULA agreement version: v1beta
Reproducibility Checklist: I certify all co-authors of this work have read and completed the Reproducibility Checklist.
Submission Number: 15286

Decision ✕ Official Review ✕

Filter by author...

Search keywords...

Sort: Newest First

Everyone | Program Chairs | Submission15286... | Submission15286... | Submission15286... | Submission15286... 5 / 11 replies shown

Add: **Withdrawal**

Paper Decision

Decision by Program Chairs 09 Dec 2024, 17:14 (modified: 09 Dec 2024, 18:15)

Program Chairs, Area Chairs, Senior Program Committee, Authors Revisions (/revisions?id=0o5HLHFvG3)

Decision: Reject

Comment:

This paper proposes a unified, hierarchical meta-representation of musical structure (STG). It leverages STG to develop a measure of structural distance and derives a centroid STG representative of a music corpus. The reviewers raised concerns about the lack of details to assess experimental results and (maybe) overstatement of the method. We acknowledge the authors' concern regarding one review that is suspected of being generated by AI. In our final decision, we did not include that particular review in question and its scores in our consideration. However, considering other valid reviews, the paper is still not qualified for acceptance.

Review of "Synthesizing Composite Hierarchical Structures from Polyphonic Music Corpora Using Structural Temporal Graphs and Simulated Annealing"

Official Review by Program Committee 9qKG 01 Nov 2024, 11:41 (modified: 09 Dec 2024, 17:50)

Program Chairs, Area Chairs, Senior Program Committee, Program Committee, Authors

Revisions (/revisions?id=IXPXF1kzZ6)

Review:

This paper presents a novel method for synthesizing composite hierarchical structures from polyphonic music corpora. The authors introduce the Structural Temporal Graph (STG) as a unified meta-representation of musical structure. An STG is a k-partite directed acyclic graph where the levels form the structural hierarchy of the music, and the edges convey temporal relationships between adjacent levels. To measure the structural distance between two STGs, the authors use a simulated annealing algorithm based on graph isomorphism. Finally, to extract a representative central STG from a music corpus, the authors formulate the problem as a double NP-hard combinatorial optimization problem and propose a method that combines nested simulated annealing with an SMT solver. Experimental results show that the structural distance can accurately distinguish different musical pieces, and the computed central STG effectively summarizes the overall structure of the entire music corpus.

Strengths Innovativeness: The proposed STG model and structural distance measurement method are innovative. Methodology: The authors formalize the problem of summarizing musical structure as a double NP-hard combinatorial optimization problem and solve it using a combination of nested simulated annealing and an SMT solver. Application Potential: The method proposed in this paper has broad application prospects in music analysis and may also play an important role in areas such as music composition and music education.

Weaknesses Intuitive Problem Presentation: The paper could benefit from a more intuitive presentation of the problems identified, particularly the motivation for introducing the STG. Related Work Analysis: A more detailed analysis of the related work would help better position the contributions of this paper. Organization of Method Description: The current version of the method description is somewhat disorganized. It is recommended to optimize the organization to make it clearer and more understandable. Technical Details: More technical details about the implementation of the simulated annealing algorithm, such as parameter selection and cooling schedule, should be provided to enable readers to reproduce the experimental results. Comparison Experiments: While the experimental design is reasonable, more comparative experiments, such as those against existing methods, are suggested to further validate the superiority of the proposed method. Ablation Studies: More ablation studies are recommended to demonstrate the effectiveness of each component of the proposed method. Practical Applications: Additional practical application cases could be included in the experimental section to showcase the potential value of STG in areas like music composition and analysis. Discussion of Limitations: The conclusion section should include a more thorough discussion of the limitations of the proposed method and potential directions for future improvement. Review of Related Work: The introduction could briefly review related work to better position the contributions of this paper. For example, some existing methods for music structure analysis and their limitations could be mentioned. The STG model's high computational demands, due to NP-hard optimization and resource-intensive techniques like simulated annealing (SA) and SMT solvers, limit its scalability. This complexity poses challenges for application on larger datasets or in resource-constrained environments, as seen with GPU memory constraints for centroids of more than nine STGs. To improve applicability, consider discussing possible optimizations, such as approximate algorithms or data reduction methods, which could enhance the model's usability across various datasets. The choice of a k-partite Directed Acyclic Graph (DAG) for STG would benefit from further justification. For example, alternative graph models like undirected graphs or multigraphs could potentially offer advantages for representing non-linear or repeating musical structures. Additionally, a comparison with other graph-based approaches in music analysis (e.g., topographic maps or formal grammars) would better position the unique strengths of STG and clarify its suitability for polyphonic music data. Ablation Studies: While the introduction of the structural temporal graph (STG) highlights a hierarchical structure of finer musical features and their temporal relationships, it lacks empirical validation in the experiments. Details: The figures and tables in the article should be made more coherent and visually appealing. It is recommended that the authors review additional relevant literature to enrich the design and presentation of these graphical elements, ensuring they effectively communicate the findings and concepts.

Rating: 2: Strong rejection
Confidence: 4: The reviewer is confident but not absolutely certain that the evaluation is correct

Interesting, but the experiments are limited in scope

Official Review by Program Committee m2Wg 26 Oct 2024, 20:32 (modified: 09 Dec 2024, 17:50)
Program Chairs, Area Chairs, Senior Program Committee, Program Committee, Authors
Revisions (/revisions?id=WuYhkO0P2S)

Review:
The paper studies the problem of using multi-layer analysis of classical piano music composition and directed-graph isomorphism to classify the compositions. The principal idea is to introduce multiple concepts for decomposing the composition (using MIDI data and performance/rendering) and encode the decompositions into a directed graph. Once each composition is represented as a graph, one can use graph algorithms to compute the similarity between two compositions and then compute the centroid. Since the number of nodes is somewhat linear to the number of nodes, normalizing the graphs for comparison is an important issue. Specifically, when two graphs subject to comparison have an unequal number of nodes, the paper suggests padding the smaller graph with additional nodes. Computing the best alignment between two graphs is NP-complete, so one cannot hope to compute the best align the input graphs in a reasonable amount of time when the graphs are large. The paper proposes to use Simulated Annealing to solve the alignment problem. The authors appear to be successful in configuring SA for the purpose. The lack of details of the pieces used makes it difficult to assess the results when it comes to experimental verification. In addition, the use of different sets of composers for the two experiments is confusing. The second set of experiments shows that the centroid calculation is successful, but you also see that the pieces used for computing the centroid are very close to each other.

An important issue is that the distance is oblivious to the order in which subjects appear. In other words, if a piece has subject A and then subject B, and another piece has subject B and then subject A, the distance is very small. This order oblivious nature will be more prominent when the number of subjects in the pieces is large.

Overall, while I think that the use of SA is interesting, the exploration can be more carefully done.

Rating: 5: Marginally below acceptance threshold
Confidence: 5: The reviewer is absolutely certain that the evaluation is correct and very familiar with the relevant literature

A novel approach describing symbolic music similarity

Official Review by Program Committee ZgjR 22 Sept 2024, 22:47 (modified: 09 Dec 2024, 17:50)

Program Chairs, Area Chairs, Senior Program Committee, Program Committee, Authors

Revisions (/revisions?id=5NoHwpZ82C)

Review:
This paper presents a symbolic music representation based on (tree-like) directed graphs. The graph encodes the hierarchical structure of music, including section, motif, chord, melody, etc. Musical similarity can therefore be derived from the graph distance (a dual NP hard problem). In general, this paper is well written. The first three sections are especially good, showing that the authors are knowledgeable in this field. The experiments basically validated the proposed distance metric, though it is not presented with real-world tasks (this is a weak part of this paper).

Questions:

- The authors stated that the previous models (e.g., GTTM) are not fully automatic, but I don't think this is an adequate argument. It should be noted that the authors utilized a lot of MIR/ music processing methods to automize their representation, so, I think the same strategy can also be applied to GTTM? In other words, once you have trained those symbolic music recognition models, they can all be automated.

Rating: 7: Good paper, accept
Confidence: 4: The reviewer is confident but not absolutely certain that the evaluation is correct

A very ambitious paper, and the arguments become even stronger with the evidence presented to show the importance of hierarchical integration.

Official Review by Program Committee RuMq 19 Sept 2024, 22:24 (modified: 09 Dec 2024, 17:50)

Program Chairs, Area Chairs, Senior Program Committee, Program Committee, Authors

Revisions (/revisions?id=9AmpHc762h)

Review:
This paper presents the first method for organizing and representing symbolic musical structures at the level of the entire music corpus. Music (especially Western music) is made up of various elements that interact hierarchically, such as harmony, rhythm, motifs, and repetitive structures. In order to express the hierarchical structure of music in this way, the authors propose a method of converting existing methods into a graph structure called a structural temporal graph, which is used as a module to capture the structure at various levels. Thanks to the fact that each piece of music can be represented as a graph, it is possible to explicitly capture the concepts of centroid and structural summary of typical music corpus, for example. This is a very hard problem in terms of computational complexity, but the authors have shown a practical solution using an annealing method. The proposed method has been demonstrated by applying it to existing high-quality music corpus.

Strengths:

- This paper deals with a very ambitious theme, and seems to be the first work to have practically captured the structure at the music corpus level.
- The introduction of the problem of finding the centroid of a collection of STGs is very important in terms of the novelty and significance of the problem setting, and the authors also present a reasonable solution. More specifically, I think that the topics in Section 5 are all excellent, both in terms of their significance and importance, the interest of the issues themselves, and the practical solutions the authors propose.

Potential weaknesses:

While the overall theme is very ambitious and very attractive, there are some concerns about the detailed technical aspects.

[Major]

Lack of experimental evidence for the importance of utilizing clues from all levels

As mentioned in Section 2.2, while there have been attempts to represent the various levels of musical structure in a unified way, it is clear that there is still no method that can handle all the levels that can be considered at the moment, but in terms of corpus-level organization and expression contexts, I feel that the evidence for how important it is to actually handle all the levels in a practical and empirical way is somewhat lacking. For example, if the authors could show experimentally how much performance is reduced when using STG with some of the hierarchy clues removed, it might help readers to understand.

Concerns about the consistency and uniqueness of the hierarchical diagram and STG

I have concerns about the parts I have not been able to follow from Section 3.2 to Section 4.1. In creating a graph STG, I understand that the practical aim is (1) to have each individual algorithm at each level perform analysis, and (2) to automatically integrate them into a single STG. At this point, given the analysis results of the individual algorithms of each layer, is the corresponding STG uniquely determined? As mentioned in the note in the upper right of page 3, the inconsistency between Figure 3 and Figure 2 is explained as an effect of computer-generated generation. However, I don't understand whether the STG is determined to be in the first place from Figure 2 in the first place. For example, I am concerned that it may not be possible to establish a one-to-one correspondence (or surjection) between diagrams like Figure 2 and the STG in Figure 3 without adding restrictions such as the degree limit for each hierarchical node. For this reason, I feel that the motivation for introducing data expansion in Section 4.1 is somewhat heuristic. I am aware that it is possible that I am missing or misunderstanding an important point.

[Minor]

Although I can imagine that this is not the part that the authors wanted to focus on the most in this paper, I still get the impression that it is a bit of a lost opportunity to apply the analysis algorithm for each layer separately. For the analysis of a given layer, we can imagine that both top-down clues from the layer above and bottom-up clues from the layer below play an important role. Therefore, I imagine that it would be more effective to analyze the whole in a unified way than to analyze each layer separately. Again, I think this is not the main point of this paper, so this is just a minor comment. If STG is a robust integration method that essentially avoids this point, I think it would be a very good method.

Typo: First sentence of Section 4.1 - two STGs is is

Rating: 5: Marginally below acceptance threshold

Confidence: 3: The reviewer is fairly confident that the evaluation is correct