GRAPHDITTY: A SOFTWARE SUITE FOR GEOMETRIC MUSIC STRUCTURE VISUALIZATION

Christopher J. Tralie
Duke University Department of Mathematics

ABSTRACT

In this work, we present a new twist on music structure analysis and visualization. We devise a technique [1] to create clean audio self-similarity matrices at the song level by fusing multiple features upstream. We then derive multiple geometric features from this representation to elucidate hierarchical structure, including Laplacian eigenvectors, spring graph layouts, and diffusion maps. We then provide a suite of Javascript visualization tools to view the SSMs and derived features synchronized with the audio they represent. Our code is clean with the help of Numpy/Scipy/Librosa on the Python end and d3.js on the Javascript end, but it can be treated as a blackbox for users who would like to engage with the visualizations without delving into the technical details. Code can be found at `http://www.github.com/ctralie/GraphDitty`, and a live demo is present at `http://www.covers1000.net/GraphDitty`.

1. SIMILARITY FUSION

The self-similarity matrix (SSM) is a common data structure through which to visualize recurrence in musical audio. For a particular feature type, the SSM is a symmetric distance matrix D which records all pairwise distances between windows in time, as measured by that feature. Let D^C be a matrix measuring the cosine distance between stacked-delayed [2] [7] chroma features, and let D^M be a matrix measuring the Euclidean distance between stack-delayed MFCCs. The we can apply a *similarity kernel* to each of them so that $W_{ij}^C = \exp(-(D_{ij}^C)^2/(2\sigma_{ij}^2))$, and likewise for W^M for D^M, where σ_{ij} is a mutual nearest neighbor autotuned distance (see [11, 12] for more details); that is, large values indicate more similar windows. We then run a graph-based algorithm known as *similarity network fusion (SNF)* [3, 11, 12] to create an aggregated similarity kernel W^F from W^C and W^M, which promotes the strengths of both feature types and mitigates their weaknesses. This is similar to what we did for cover songs in [8], though it works on self-similarity instead of cross-similarity, and it does not require beat-synchronous features. We can also compute eigenvectors of the graph laplacian on W^F indicator functions of hierarchical structural elements, as in [6]. This combination of stacked delay embeddings and SNF can be viewed as a more general, global alternative to similarity diagonal promotion which has previously been used to preprocess the graph Laplacian [6]. As can be seen in Figures 1 and 2, it at least qualitatively does a much better job at making clean similarity matrices and Laplacian eigenvectors than Chroma by itself.

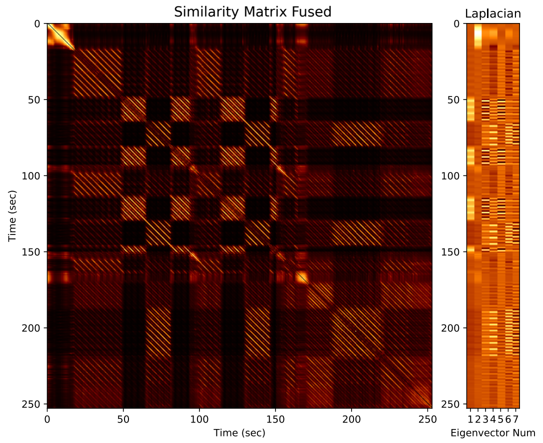

Figure 1. Similarity matrix W^F and Laplacian eigenvectors after applying similarity network fusion to stack-delayed Chroma and MFCC features for the song “Smooth Criminal” by Michael Jackson. The SSM and eigenvectors are much cleaner than those with just raw chroma in Figure 2.

2. VISUALIZING TIME-ORDERED SIMILARITY

The first facet of our GUI simply allows the user to view audio synchronized with the SSM, but enables a powerful way to visualize and jump between repeated elements in the song [3]. The second visualization performs a spring layout of the weighted graph induced from W^F, with the help of d3.js [2], as shown in Figure 3. Since we have applied a similarity kernel, the spring constant is proportional to how similar windows are; the simulation encourages more

[1] This is a refinement of / followup to our prior works “scaffolding and spines” [1] and “Loop Ditty” [9]

[2] We use stack-delayed features to promote diagonal structures, as shown in [10]

[3] This is similar to the cross-similarity GUI viewer we created for cover songs [8].

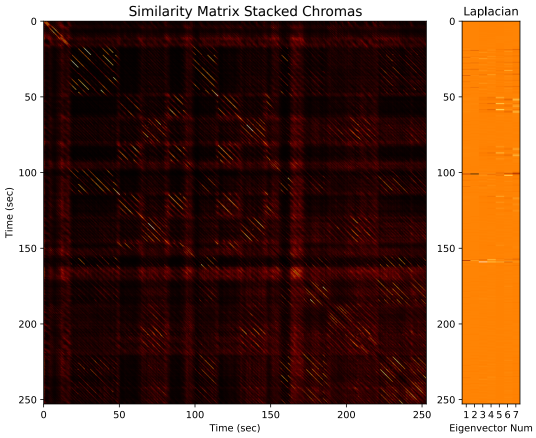

Figure 2. Similarity matrix W^C using the cosine distance on stack-delayed Chroma features, along with the corresponding weighted Laplacian eigenvectors. While the stacked delay embedding helps diagonals to appear which indicate repeated structure, it is not as clean as the fused SSM in Figure 2.

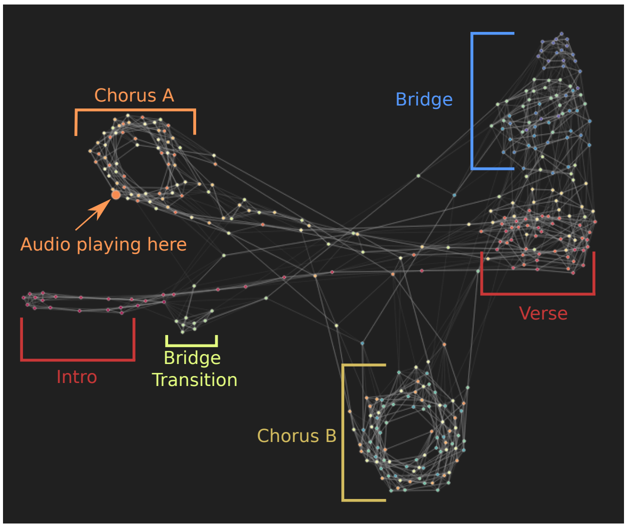

Figure 3. A dynamic weighted spring layout based on the weights in Figure 1, which is rendered with the help of d3.js [2].

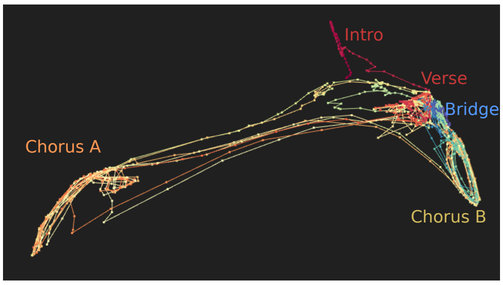

Figure 4. 3D Diffusion maps rendered by WebGL, synchronized to the music.

similar windows to be closer together. Note that the simulation is dynamic; nodes in the graph can be moved around, and the simulation will settle in a local min of energy. Finally, we present a GUI for showing music synchronized to 3D diffusion maps [4]. This is the most similar GUI to our previous "Loop Ditty" GUI [9], though it works purely on similarity information and not on coordinates in feature space, so it is much more general.

In future work, we would like to explore all of these structures for pruning in large scale audio cover song identification, similar to the aligned hierarchies work [5] on symbolic cover song identification.

3. REFERENCES

[1] Paul Bendich, Ellen Gasparovic, John Harer, and Christopher Tralie. Geometric models for musical audio data. In *Proceedings of the 32st International Symposium on Computational Geometry (SOCG)*, 2016.

[2] Michael Bostock et al. D3. js. *Data Driven Documents*, 492:701, 2012.

[3] Ning Chen, Wei Li, and Haidong Xiao. Fusing similarity functions for cover song identification. *Multimedia Tools and Applications*, pages 1–24, 2017.

[4] Ronald R Coifman and Stéphane Lafon. Diffusion maps. *Applied and computational harmonic analysis*, 21(1):5–30, 2006.

[5] Katherine M Kinnaird. Aligned hierarchies: A multi-scale structure-based representation for music-based data streams. In *16th International Society for Music Information Retrieval (ISMIR)*, pages 337–343, 2016.

[6] Brian McFee and Daniel PW Ellis. Analyzing song structure with spectral clustering. In *15th International Society for Music Information Retrieval (ISMIR)*, 2014.

[7] Joan Serra, Xavier Serra, and Ralph G Andrzejak. Cross recurrence quantification for cover song identification. *New Journal of Physics*, 11(9):093017, 2009.

[8] Christopher J Tralie. Early mfcc and hpcp fusion for robust cover song identification. In *18th International Society for Music Information Retrieval (ISMIR)*, 2017.

[9] Christopher J Tralie. *Geometric Multimedia Time Series*. Duke ph.d. dissertation, Department of Electrical and Computer Engineering, Duke University, 2017.

[10] Christopher J. Tralie and Jose A. Perea. (quasi)periodicity quantification in video data, using topology. *SIAM Journal on Imaging Sciences*, 11(2):1049–1077, 2018.

[11] Bo Wang, Jiayan Jiang, Wei Wang, Zhi-Hua Zhou, and Zhuowen Tu. Unsupervised metric fusion by cross diffusion. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2997–3004. IEEE, 2012.

[12] Bo Wang, Aziz M Mezlini, Feyyaz Demir, Marc Fiume, Zhuowen Tu, Michael Brudno, Benjamin Haibe-Kains, and Anna Goldenberg. Similarity network fusion for aggregating data types on a genomic scale. *Nature methods*, 11(3):333–337, 2014.