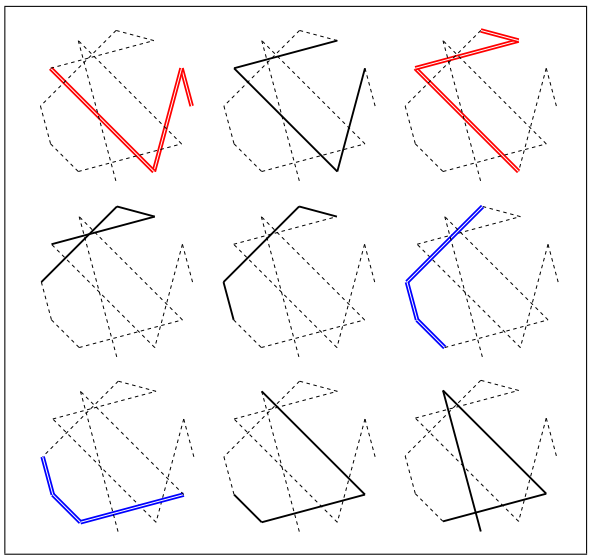

Figure 3. Clock diagrams of the tone row $(0, 1, 10, 5, 2, 3, 6, 7, 8, 11, 4, 9)$ showing the nine sequences of length 4. The two recurring motives are indicated with doubled lines.

repeated eleven times to form the row. In the opposite case, the *all-interval rows* consist of every different musical interval and consequently possess very few motives. Figure 4 shows clock diagrams for the circle of fifths (containing the maximum number of motives of all lengths) and an all-interval row that contains no motives above length 2 at all.

Our approach has two stages: first, we enumerate the motives of all lengths for every row class; then, we advance four plausible models for assigning an overall symmetry rating to any set of motives, with attention to the music-theoretic and mathematical literature to ensure consistency of our results with existing knowledge. We begin the first stage by treating each tone row as a collection of length-d sequences of notes, with d ranging from 2 to 12. Figure 3 shows a particular tone row's nine sequences of length 4, of which two pairs are found to be separate recurring motives. In general, the set of length-d motives of a tone row can be quantified by a partition of $12 - d + 1$ (the total number of length-d sequences) based on which sequences appear multiple times. Figure 3 shows that the depicted tone row corresponds to the partition $9 = 2+2+1+1+1+1+1$ since there are two motives that each appear twice, plus five other length-4 sequences that do not form motives. As d varies from 2 to 12, the corresponding sets of motives fully encode the symmetrical properties of a tone row at every cardinality. Therefore, the symmetrical properties of any tone row can be represented by a list of partitions of the integers from 1 to 11, corresponding to values of d ranging from 12 down to 2. We refer to such a list of partitions as the *full symmetry data* of a row class.

Generalizing symmetry in this way encodes a tremendous amount of information. Hunter and von Hippel identify three types of rows (eight, when cyclic shift is included) [8]. Friptertinger and Lackner extend this to a few thousand by introducing tropes [10]. But a naïve upper bound on the number of distinct full symmetry data is the product of the first eleven partition numbers, about 5.379×10^{10} or five thousand times the number of row classes! There are many principled ways to simplify the full symmetry data while maintaining consistency with the literature. One promising approach (not explored in this paper) would map each partition to its length, summarizing the symmetry data of a tone row with an integer 11-tuple. Under this scheme, rows whose 11-tuple values are small exhibit more repetitive motives (hence, greater symmetry) than those whose 11-tuple values are large. In an extreme example, the chromatic scale and the circle of fifths achieve scores of all 1's in this scheme. In this paper, we choose to map the full symmetry data to a differently defined 11-tuple: the maximum values of the partitions. The reason for using the maximum value instead of the length is subjective as both schemes capture essential information about the partition. Figure 5 depicts a tone row whose most-repeated motives of length 2, 3, and 4 have multiplicities of 3, 4, and 2, respectively. This particular tone row has no motives of length 5 or greater, meaning that its symmetries are quantified by the tuple $(3, 4, 2, 1, 1, 1, 1, 1, 1, 1, 1)$. We refer to this tuple as the *symmetry score* of the row class.

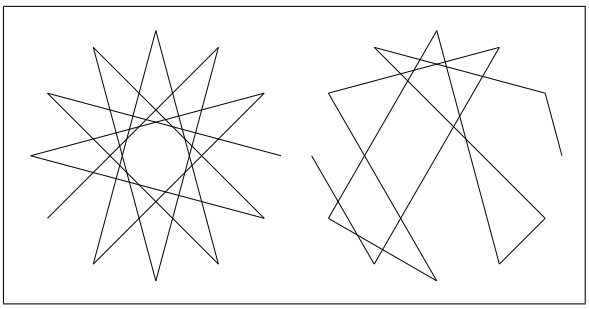

Figure 4. Clock diagrams of the maximally symmetric circle of fifths and a minimally symmetric all-interval row.

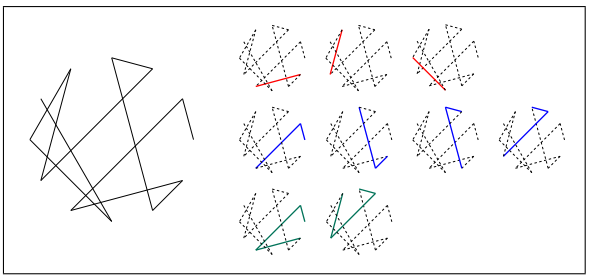

Figure 5. Clock diagrams of the tone row $(0, 1, 8, 11, 10, 3, 2, 7, 4, 6, 9, 5)$ showing that it achieves symmetry scores of 3, 4, and 2 for the lengths of 2, 3, and 4, respectively.

3. ANALYSIS

In Section 2 we established a map from the set of row classes to an 11-dimensional lattice with the property that greater values correspond to higher multiplicity of motives. The two clock diagrams in Figure 4 illustrate this property:

whereas the circle of fifths maps to a symmetry score of $(11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1)$, the depicted all-interval row maps to a symmetry score of $(2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$ which is in fact the lowest possible multiplicity of motives among all row classes. The next step is to assess whether the Viennese tone rows contain unusually high levels of symmetry, but this depends on how we compare symmetry scores to each other. Instead of presenting just one possible interpretation of "high symmetry," we give four reasonable candidates that span a wide range of interpretations of symmetry while still respecting musical intuition and the mathematical literature on the topic. In each model, we conduct hypergeometric tests to gauge whether the incidence of high symmetry in the composers' corpora was significantly greater than would be predicted by chance alone. While these tests are exploratory in nature, the models in Sections 3.1 and 3.2 admit a conservative Bonferroni correction to adjust for the issue of multiple testing [14].

3.1 Reverse-Lexicographic Model

The most direct extension of Hunter and von Hippel's results is to introduce a sensible total order on the set of symmetry scores, then split the scores into "low" and "high" bins in music-theoretically sound ways and apply a hypergeometric test. A natural candidate for the total order is reverse-lexicographic order (RLEX), prioritizing longer motives over shorter ones. This choice has the advantage of seamlessly reproducing known results in its base case; moreover, it yields a natural definition of low and high symmetry: for each length d, the valid splits are the locations in the total order at which the d-symmetry score increments. For example, we can split the set of tone rows according to whether any length-4 motive is present; in the set of all row classes, 23.62% have a length-4 motive, yet fully 50% of Schoenberg's tone rows exhibit this symmetry. The associated hypergeometric test gives a p-value of $p < 0.0002$, so the prevalence of motives of length at least 4 in Schoenberg's work is unlikely to have arisen by chance alone. This model has 38 splits (40 for Berg) and the adjusted p-values are computed by multiplying by these numbers. Table 1 shows a selection of significant results for the three composers. Notably, more than 90% of Webern's corpus exhibits a statistically significant level of symmetry in the RLEX sense and both Schoenberg and Webern's corpora exhibit significant levels of symmetry under the multiple testing correction. Berg's use of a fourth symmetry and the correspondingly smaller space of row classes makes it more difficult to obtain significant results for his corpus, as already noted by Hunter and von Hippel. However, even in this case a few notable results are seen, although only on an exploratory basis.

3.2 Lattice Rank Model

The RLEX model faithfully reproduces known results, yet strongly penalizes row classes that contain many short motives but few long ones. The Lattice Rank model delivers on the opposite extreme by using a grading according to the sum of the symmetry score, thereby equally weighting motives of all lengths. Here, the natural divisions of "high" and "low" symmetry are determined by the rank function with 54 distinct levels (109 for Berg). For example, two-thirds of Webern's tone rows have lattice rank of 6 or greater (that is, 6 more multiplicity of motives than the lowest symmetry score), while only 11.17% of the set of all row classes has this property. The quantity of tone rows in Webern's corpus with at least this lattice rank almost certainly did not arise by chance alone ($p = 2.5 \times 10^{-9}$). Table 2 shows that the Lattice Rank model obtains remarkably similar results to the RLEX model in spite of being agnostic to motive length. Again, Schoenberg and Webern's corpora exhibit statistically significant levels of symmetry even under the multiple testing correction, while Berg's corpus only does so on an exploratory basis.

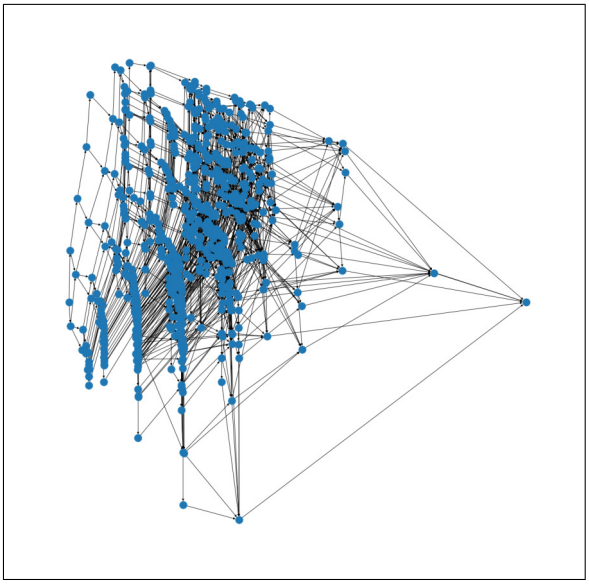

Figure 6. Directed acyclic graph representation for the partially ordered set of symmetry scores relevant to Schoenberg and Webern.

The fact that the RLEX and Lattice Rank models yield similar results in spite of applying diametrically opposed weighting schemes to motives of different lengths suggests that the Viennese tone rows simply contain a lot of internal symmetries no matter how you count them. Still, these models impose value judgments on the relative importance of each internal symmetry, possibly affecting the outcome of the analysis. To validate our findings, we offer a non-parametric alternative that makes no assumptions at all about the relative importance of the different motives. Instead, this alternative is based on the single intuitive assumption that the symmetry scores form a partially ordered set: a row class whose symmetry score is no greater in any component than the symmetry score of a different row class cannot be said to exhibit greater symmetry overall. Abstracting from the lattice structure in this way results in the diagrams in Figures 6 and 7, where each vertex is a symmetry score and the directed edges (all running left-to-right) run from lesser to greater in the partial order.

Schoenberg	All	p-value	Webern	All	p-value	Berg	All	p-value
76.19%**	45.50%	5.11×10^{-5}	90.48%***	39.15%	1.51×10^{-6}			
50.00%**	23.62%	1.88×10^{-4}	57.14%*	23.68%	1.01×10^{-3}	47.83%	28.84%	0.042
38.10%***	11.11%	5.47×10^{-6}	42.86%**	11.16%	2.23×10^{-4}			
11.90%	2.29%	2.67×10^{-3}	28.57%***	2.29%	5.89×10^{-6}	8.70%	1.04%	0.024
4.76%	0.13%	1.50×10^{-3}	19.05%***	0.13%	1.93×10^{-8}			

Table 1. Selected RLEX results for each composer, comparing their corpora to the set of all row classes with respect to various "high" and "low" symmetry bins. Asterisks indicate significance level under conservative Bonferroni correction. * : $p < .05$, ** : $p < .01$, *** : $p < 0.001$.

Schoenberg	All	p-value	Webern	All	p-value	Berg	All	p-value
83.33%	66.07%	0.011	95.24%	66.07%	1.96×10^{-3}			
47.62%**	20.87%	9.98×10^{-5}	71.43%***	20.87%	9.13×10^{-7}	56.52%	35.04%	0.029
33.33%**	11.17%	1.17×10^{-4}	66.67%***	11.17%	2.54×10^{-9}			
9.52%*	0.90%	5.48×10^{-4}	38.10%***	0.90%	7.48×10^{-12}	17.39%	4.50%	0.018
7.14%***	0.11%	1.47×10^{-5}	19.05%***	0.04%	1.38×10^{-10}			

Table 2. Selected Lattice Rank results for each composer, comparing their corpora to the set of all row classes with respect to various "high" and "low" symmetry bins. Asterisks indicate significance level under conservative Bonferroni correction. * : $p < .05$, ** : $p < .01$, *** : $p < 0.001$.

3.3 Partial Order Rating Models

In order to set up a hypergeometric test on the poset of symmetry scores, we still need to choose some linear extension of the poset to a total order. The RLEX and Lattice Rank models accomplish this by appealing to music intuitions on the relative importance of different kinds of motives. However, the fact that our poset has a unique least and greatest element gives us a non-parametric alternative. Given any symmetry score v, we count the number of row classes with strictly lesser scores (ascendants) and strictly greater scores (descendants) in the poset. If v has relatively few descendants compared to ascendants, then few row classes have strictly greater symmetry than v and we may say that v possesses a relatively high level of symmetry. To make this idea precise, let A_v be the number of row classes assigned to strictly lesser scores than v, and D_v be the number of row classes assigned to strictly greater scores than v. Lastly, let M be the total number of row classes. The formula

$$\frac{1}{2}(1 + (A_v - D_v)/M))$$

ranges from 0 to 1 as v ranges from the least to greatest elements of the poset. We define the Solo Poset Rating of v by renormalizing this formula to range from 0 to infinity, like so:

$$\begin{aligned} PR_s(v) &= -\log\left(1 - \frac{1}{2}(1 + (A_v - D_v)/M))\right) \\ &= -\log\left(\frac{1}{2}(1 - (A_v - D_v)/M)\right). \end{aligned} \quad (1)$$

In Figure 7, the symmetry score marked by the black vertex would receive a relatively high Solo Poset Rating because it has many ascendants but relatively few descendants. This imbalance is further amplified by the fact that scores near the minimum symmetry score are generally assigned to many more row classes than scores near the maximum symmetry score.

For example, two of Berg's tone rows have a symmetry score of $(6, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1)$, whose Solo Poset Rating is 0.695. An additional ten of Berg's tone rows achieve a greater Solo Poset Rating than this, so $12/23$ or 52.17% of Berg's tone rows exhibit a Solo Poset Rating of 0.695 or greater. Meanwhile, just 29.42% of all tone rows achieve a Solo Poset Rating of at least 0.695. If Berg had selected row classes at random, the probability that twelve or more of his corpus would have a Solo Poset Rating of at least 0.695 is just 0.018.

Figure 7 shows that for any given symmetry score, there are many other scores in the poset that are incomparable. We define the Cohort Poset Rating of a particular score as the average of the Solo Poset Ratings of all scores that are incomparable with the chosen score. Both variants of the Poset Rating quantify how close a given score is to the highest possible symmetry score, compared to its closeness to the lowest possible symmetry score. The two ratings yield nearly identical results when analyzing the corpora of Schoenberg and Webern (Tables 3 and 4). Due to Berg's smaller space of row classes, the only noteworthy finding for his corpus under the Poset Rating models is the one given in the previous paragraph (the Solo and Cohort models coincide in this case).

4. DISCUSSION

Our work takes a flexible approach to detecting patterns within tone rows, uncovering partial symmetries that cannot be detected by group theory. The trade-off is that we also find many internal structures that have no name in music theory. Whether this is a true disadvantage depends on

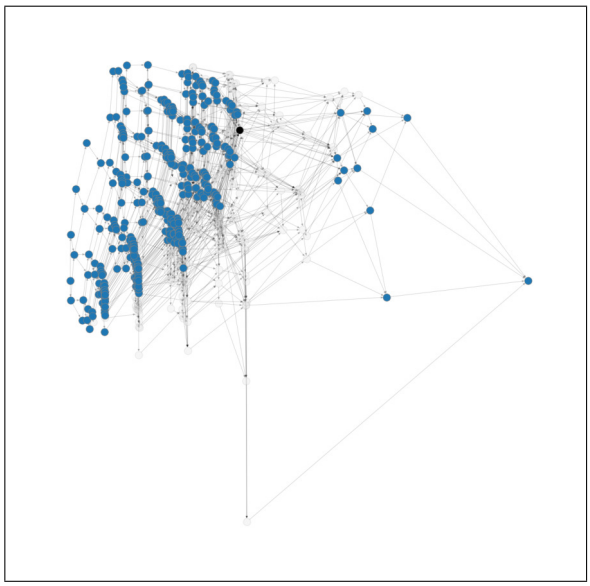

Figure 7. Directed acyclic graph representation of the symmetry scores relevant to Berg with highlighted ascendants and descendants of the black vertex, and incomparable scores greyed out.

Schoenberg	All Solo (Cohort)	p-value
83.33%	65.87%	0.010
52.38%	26.23%	2.73×10^{-4}
33.33%	12.85% (11.65%)	5.13×10^{-4}
9.52%	0.74% (0.60%)	2.73×10^{-4}
4.76%	0.74% (0.13%)	9.08×10^{-3}

Table 3. Selected Poset Rating results for Schoenberg, comparing his corpus to the set of all row classes with respect to various "high" and "low" symmetry bins. When the Solo and Cohort ratings do not coincide, the larger of the two p-values is displayed for brevity.

Webern	All Solo (Cohort)	p-value
95.24%	45.35%	1.62×10^{-6}
66.67%	12.85% (16.03%)	2.78×10^{-7}
52.38%	2.85%	2.67×10^{-12}
33.33%	1.03% (1.06%)	1.51×10^{-9}
19.05%	0.12% (0.25%)	2.39×10^{-7}

Table 4. Selected Poset Rating results for Webern, comparing his corpus to the set of all row classes with respect to various "high" and "low" symmetry bins. When the Solo and Cohort ratings do not coincide, the larger of the two p-values is displayed for brevity.

the validity of our basic premise: that the presence of a recurring motive is noteworthy in itself. We believe that this premise is supported by the preexisting scholarly interest in derived rows, those tone rows that are fully generated by a shorter repeating motive [10]. Since small adjustments to a derived row result in similar rows that yet lack any official symmetry, it seems natural to extend the concept of derived row to any tone row that is at least partially generated by a motive.

In bringing new methods to this topic, we have made particular effort to situate our work in the literature and reproduce known results where possible. Indeed, all our models agree with the existing knowledge that roughly 5% of Schoenberg's corpus and 19% of Webern's corpus have significant incidence of symmetry at the highest level. The RLEX model in particular achieves exact replication of known results. Where it was necessary for us to make a value judgment on the relative importance of different motives, we have followed the literature (in the case of RLEX) and also explored the diametrically opposed value judgment (in the case of Lattice Rank) in order to provide maximum contrast. In each case, we performed a multiple testing correction and still found significant levels of symmetry in much of the tone rows of Schoenberg and Webern. We also developed two non-parametric models to further validate our findings. Moreover, our approach to defining symmetry by the presence of shorter motives is analogous to von Hippel and Huron's measurement of the "tonalness" a row by the tonal properties of its subsets [7].

As for whether the Viennese twelve-tone composers favored symmetry in their work, our models support this across the board. We find that all three composers made extensive use of symmetry in their work, decidedly beyond what can be explained by chance alone. We found that 76%–83% of Schoenberg's corpus and 90%–95% of Webern's corpus contained significant levels of symmetry (up from 5% and 19%, respectively.) Even in the case of Berg, whose much smaller space of row classes limits the power of the hypergeometric test, we can extend the previously determined figure of 9% to the range of 48%–57% of his corpus although only on an exploratory basis.

Our combinatorial approach has one other advantage over the group-theoretic approaches in the literature. By forgetting the permutation group structure and studying motives on a purely combinatorial level, we no longer restrict ourselves to studying tone rows. Indeed, our methods apply just as well to any data series together with a relevant group of symmetries, with the one concession that there is no clear analog for the hypergeometric test. Still, the n-tuple symmetry score can serve as a feature for classification of melody and rhythm that encodes information of many different cardinalities, evoking concepts from persistent homology and wavelet analysis. While we assert no formal connection between these fields and our work, we take inspiration from the concept of studying an object at many different scales and hope that our success in the enumeration of tone row motives may find broader application in music information retrieval.

5. REFERENCES

[1] E. Haimo, "The evolution of the twelve-tone method," in *The Arnold Schoenberg Companion* (W. B. Bailey, ed.), pp. 101–128, Westport, CT: Greenwood Press, 1998

[2] K. Bailey, *The Twelve-Note Music of Anton Webern: Old Forms in a New Language*. Cambridge: Cambridge University Press, 1991.

[3] D. Headlam, *The Music of Alban Berg*. New Haven, CT: Yale University Press, 1996.

[4] J. M. Hauer, *Zwölftontechnik : die Lehre von den Tropen*. Wien: Universal Edition, 1925.

[5] E. Amiot, "Mathématiques et analyse musicale: une fécondation réciproque," in *Analyse Musicale*, vol. 28, pp. 37–41, 1992.

[6] E. Krenek, "Musik und Mathematik," in *Über neue Musik*, pp. 71–89, Vienna, Verlag der Ringbuchhandlung, 1937

[7] P. T. von Hippel and D. Huron, "Tonal and 'Anti-Tonal' Cognitive Structure in Viennese Twelve-Tone Rows," in *Empirical Musicology Review*, vol. 15, No. 1-2, 2020.

[8] D. J. Hunter and P. T. von Hippel, "How Rare is Symmetry in Musical 12-Tone Rows?," in *The American Mathematical Monthly*, vol. 110, No. 2, pp. 124–132, February 2003.

[9] M. Babbitt, "Some aspects of twelve-tone composition," in *The Score and I.M.A. Magazine*, vol. 12, pp. 53–61, 1955.

[10] H. Fripertinger and P. Lackner, "Tone rows and tropes," in *Journal of Mathematics and Music*, vol. 9, No. 2, pp. 111–172, 2015.

[11] M. Gotham and J. Yust, "Serial Analysis: A Digital Library of Rows in the Repertoire and their Properties, with Applications for Teaching and Research," in *DLfM '21: 8th International Conference on Digital Libraries for Musicology*, New York, NY: Association for Computing Machinery, 2021.

[12] C. Krumhansl, *Cognition foundations of musical pitch*, New York, NY: Oxford University Press, 1990.

[13] J. Yust, "Dimensions of Atonality: A Response and Extension of von Hippel and Huron (2020)," in *Empirical Musicology Review*, vol. 15, No. 1-2, pp. 119 - 119, 2020.

[14] M. Jafari and N. Ansari-Pour, "Why, When, and How to Adjust Your P Values?," in *Cell J.*, vol. 20, No. 4, 2019.

THE POWER OF DEEP WITHOUT GOING DEEP? A STUDY OF HDPGMM MUSIC REPRESENTATION LEARNING

Jaehun Kim
Delft University of Technology
Multimedia Computing Group
J.H.Kim@tudelft.nl

Cynthia C. S. Liem
Delft University of Technology
Multimedia Computing Group
C.C.S.Liem@tudelft.nl

ABSTRACT

In the previous decade, Deep Learning (DL) has proven to be one of the most effective machine learning methods to tackle a wide range of Music Information Retrieval (MIR) tasks. It offers highly expressive learning capacity that can fit any music representation needed for MIR-relevant downstream tasks. However, it has been criticized for sacrificing interpretability. On the other hand, the Bayesian nonparametric (BN) approach promises similar positive properties as DL, such as high flexibility, while being robust to overfitting and preserving interpretability. Therefore, the primary motivation of this work is to explore the potential of Bayesian nonparametric models in comparison to DL models for music representation learning. More specifically, we assess the music representation learned from the Hierarchical Dirichlet Process Gaussian Mixture Model (HDPGMM), an infinite mixture model based on the Bayesian nonparametric approach, to MIR tasks, including classification, auto-tagging, and recommendation. The experimental result suggests that the HDPGMM music representation can outperform DL representations in certain scenarios, and overall comparable.

1. INTRODUCTION

Deep learning became one of the most popular and successful methodologies to tackle a various range of MIR tasks; music classification [1], music generation [2], recommendation [3] and more. Some of the known benefits are 1) the high expressiveness towards any data structure, 2) effective ways to handle the overfitting, and finally, 3) the rapidly improving infrastructural supports, including both hardware (i.e., accelerators such as GPU) and software (i.e., deep learning frameworks). Fuelled by these benefits, DL models can learn useful representations of diverse data, which is one of the key reasons for its success [4]. On the other hand, DL models often are criticized for their lack of interpretability [5], also for applications within the MIR domain [6, 7].

According to [8, 9], interpretability of a machine learning model can be defined as the degree to which a user can understand the underlying mechanism of its operation. In this regard, DL models would be substantially less interpretable due to their complex internal structure than simpler—yet may be comparably expressive—alternatives such as hierarchical probabilistic models.

The Bayesian Nonparametric (BN) approach is such an alternative to achieving what DL does well in an interpretable way. As a nonparametric method, it can overcome underfitting by unlimited model capacity, while resisting overfitting through its Bayesian nature [10]. At the same time, the model can be interpretable, as the probabilistic model structure hypothesizes the data generation process in a relatively simple and humanly understandable manner.

In past decades, BN models have been applied in various MIR tasks. For example, they were applied to a latent representation of music audio in estimating music (self-) similarity [11, 12]. They also were employed to conduct harmonic and spectral analyses by decomposing the time-frequency representation of the music audio into a mixture of countable infinite latent components [13, 14]. These analyses also have shown to be useful for downstream tasks such as structural analysis [15] or music recommendation [16]. Further, a discriminative model based on the BN was developed for music emotion recognition [17].

While DL gained considerable popularity, it is striking BN models did not gain as much attention. As they promise similar high-level properties and strengths to DL, it will be worthwhile to explore to what extent they compare to modern DL models for MIR-relevant tasks. Therefore, in this work, we will assess music representations learned from BN models, and compare them under similar experimental control with modern deep neural network models. In particular, we consider the Hierarchical Dirichlet Process Gaussian Mixture Model (HDPGMM) [16, 18, 19] as our model of interest. To concretize the comparison, we employ the transfer learning experimental framework [20–22], which is commonly used for assessing the potential of learned representations. The contributions of this work can be listed as follows:

1. We explore and suggest insight into how "good" and transferable the HDPGMM representation is for a range of MIR tasks.
2. We provide an implementation of an efficient, GPU-

accelerated inference algorithm for HDPGMM, which can handle large-scale music datasets.

2. HDPGMM

In this section, [1] we introduce the Hierarchical Dirichlet Process Gaussian Mixture Model (HDPGMM) [12, 19]. We first discuss the Dirichlet Process Mixture Model (DPMM) upon which the HDPGMM is extended.

2.1 Dirichlet Process Mixture Model

The Dirichlet Process (DP) is an well-known stochastic process that draws random probability distributions, which often is described as a distribution over distributions [10]. One of the most common applications of DP is the infinite mixture model, thanks to the property of DP that one can draw distributions of an arbitrary dimensionality. Thus, the Dirichlet Process Mixture Model (DPMM) can find the appropriate number of components K based on the set of observations as part of the learning process.

Among several ways to represent DP, we introduce the stick-breaking construction [24] [2]. Stick-breaking constructs DP in a simple and general manner. Formally, it is as follows:

$$\begin{aligned} \beta'_k &\sim \text{Beta}(1,\gamma) & \phi_k &\sim H \\ \beta_k &= \beta'_k \prod_{l=1}^{k-1}(1-\beta'_l) & G_0 &= \sum_{k=1}^{\infty} \beta_k \delta_{\phi_k} \end{aligned} \tag{1}$$

where the two equations in the left column represent the draw of infinite dimensional weights β_k which sums to one. Notably, the distribution for β is also referred to $\beta \sim \text{GEM}(\gamma)$ [25]. In the right column, H denotes the base distribution from which variable $\phi_k \in \Phi$ is drawn. The right bottom equation defines the draw of the probability measure G_0, where δ_{ϕ_k} means the point mass centered at the component ϕ_k. Altogether, Eq. (1) constructs the DP $G_0 \sim \text{DP}(\gamma, H)$. Figure 1 depicts the process in a graphical way.

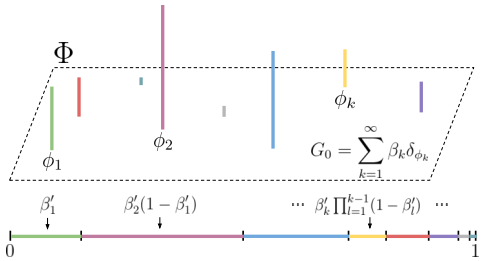

Figure 1: Illustration of stick-breaking construction.

In the mixture model context, we want to infer mixture components $\{\phi_1, \phi_2, \cdots, \phi_k, \cdots\}$ that fit to the data observations $\{x_1, x_2, \cdots, x_n\}$, which are assumed to be drawn from distributions $F(\phi)$ parameterized with variable ϕ (i.e., for a multivariate Gaussian F, $\phi = \{\mu, \Sigma\}$). We now can use DP for drawing the component ϕ that corresponds to x_i by introducing the cluster assignment variable $y_i \sim \text{Mult}(\beta)$:

$$\begin{aligned} \beta|\gamma &\sim \text{GEM}(\gamma) & \phi_k|H &\sim H \\ y_i|\beta &\sim \text{Mult}(\beta) & x_i|y_i, \{\phi_k\} &\sim F(\phi_{y_i}) \end{aligned} \tag{2}$$

where ϕ_{y_i} denotes the component parameter ϕ indexed by the assignment variable y_i corresponding to the ith observation x_i.

2.2 Hierarchical DPMM

In many data structures, groupings of atomic data points arise naturally (i.e., audio frames within a song, songs from an artist, words of lyrics). Hierarchical DP (HDP) is an extension of DP that models the "groupings" by imposing group-level DPs derived from the "corpus-level" DP as the global pool of components [19]. Following [18], the jth group-level DP can be expressed as follows:

$$\begin{aligned} \pi'_{jt} &\sim \text{Beta}(1,\alpha_0) & \psi_{jt} &\sim G_0 \\ \pi_{jt} &= \pi'_{jt} \prod_{s=1}^{t-1}(1-\pi'_{js}) & G_j &= \sum_{t=1}^{\infty} \pi_{jt} \delta_{\psi_{jt}} \end{aligned} \tag{3}$$

As seen above, HDP appears as the recursion of multiple levels of DPs [3]. Notably, the base distribution G_0 of each group-level DP is from the corpus-level DP. This relationship allows mapping group-level atoms ψ_{jt} to the corpus-level atoms ϕ_k. Wang et al. [18] introduce indicator variables $c_{jt} \sim \text{Mult}(\beta)$ which maps ψ_{jt} and ϕ_k by $\psi_{jt} = \phi_{c_{jt}}$, where β is drawn from the corpus-level DP in Eq. (1). It simplifies the model as we do not need to represent ψ_{jt} explicitly. Finally, we can represent HDPMM by introducing another indicator variable $z_{jn} \sim \text{Mult}(\pi_j)$ for the nth observation x_{jn} within the jth group, similarly to Eq. (2):

$$\begin{aligned} \pi_j|\alpha_0 &\sim \text{GEM}(\alpha_0) & \theta_{jn} &= \psi_{jz_{jn}} = \phi_{c_{jz_{jn}}} \\ z_{jn}|\pi_j &\sim \text{Mult}(\pi_j) & x_{jn}|z_{jn}, c_{jt}, \{\phi_k\} &\sim F(\theta_{jn}) \end{aligned} \tag{4}$$

where we use the indicator z_{jn} to select ψ_{jt}, which eventually is mapped as $\phi_{c_{jz_{jn}}}$ that represents the parameter θ_{jn} to draw the observation x_{jn}. HDPGMM is then defined by simply setting F as the (multivariate) Gaussian distribution and setting H as one of the distributions from which we can sample the mean and covariance (i.e., Gaussian-inverse Wishart distribution). Figure 2 depicts the HDPGMM graphically.

2.3 Inference Algorithm

We employ the online Variational Inference (VI) [18], which is a common and significantly faster choice than the Markov Chain Monte Carlo (MCMC) method to infer fully Bayesian models. VI approximates the true posterior by

[1] We adopted most of the notations from [18, 23].

[2] Literature commonly chooses the Chinese Restaurant Process (CRP) as an illustrative metaphor for DP, as it is intuitive and explains various properties of DP well [10]. We discuss DP with the stick-breaking construction, due to its further usage in the model inference within the work. Other metaphorical descriptions of DP are given in [10, 12]

[3] It implies naturally that multiple levels are possible (i.e., corpus - author - document) if it suits the data structure.