

Trade-offs between structural richness and perceptual robustness in music network representations

Lluc Bono Rosselló^{1,*}, Robert Jankowski^{2,3,4}, Hugues Bersini¹, Marián Boguñá^{3,4} and M. Ángeles Serrano^{3,4,5}¹*Institute for Interdisciplinary Studies on Artificial Intelligence (IRIDIA), Université Libre de Bruxelles, Brussels, Belgium*²*Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 2628 CD Delft, Netherlands*³*Departament de Física de la Matèria Condensada, Universitat de Barcelona, Martí i Franquès 1, E-08028 Barcelona, Spain*⁴*Universitat de Barcelona Institute of Complex Systems (UBICS), Universitat de Barcelona, Barcelona, Spain*⁵*ICREA, Passeig Lluís Companys 23, E-08010 Barcelona, Spain*

(Received 9 March 2026; accepted 8 June 2026; published 30 July 2026)

Music is a structured and perceptually rich sequence of sounds in time, whose perception is shaped by the interplay of expectation and uncertainty about what comes next. Yet, the uncertainty we infer from music depends on how the musical piece is encoded as an event sequence. In this work, we use network representations, in which event types are nodes and observed transitions are directed edges, to compare how different feature encodings shape the transition structure we recover and how robust that structure is under modeled perceptual constraints. We systematically analyze eight encodings of piano music, from single-feature vocabularies to richer multifeature combinations. These representational choices reorganize the state space and fundamentally reshape network topology, shifting how uncertainty is distributed across transitions. To connect these descriptive differences to perception, we adopt a perceptual-constraint model that captures imperfect access to transition statistics. Overall, compressed single-feature representations yield dense transition structures with higher entropy rates, corresponding to higher average uncertainty per step, yet low model error, indicating that the constrained estimate stays close to the corpus transitions. In contrast, richer multifeature representations preserve finer distinctions but expand the state space, sharpen transition profiles, lower entropy rates, and increase model error. Finally, across representations, uncertainty concentrates in diffusion-central nodes while model error remains low there, suggesting an informational landscape in which predictable flow coexists with localized surprise. Overall, our results show that feature choice shapes not only the networks we reconstruct, but also which transition statistics are available and how vulnerable they are to distortion under perceptual constraints.

DOI: [10.1103/x6hc-jb2r](https://doi.org/10.1103/x6hc-jb2r)

I. INTRODUCTION

Music is a highly complex phenomenon composed of low-level features that combine into patterns that incorporate multiple structural aspects. Gaining insight into how these patterns emerge from basic musical events offers a useful perspective for understanding both the complexity of music and the perceptual processes involved in listening [1,2]. These perceptual processes include the formation of expectations about what comes next and the experience of varying uncertainty and surprise as those expectations are confirmed or violated [2–4].

Any analysis of musical structure depends on how the musical piece is encoded as a sequence of events. Musical sequences can be represented at different levels of detail by selecting subsets of features—such as pitch, interval, or duration—that preserve some distinctions while discarding others, an idea closely related to *viewpoint*-based approaches

in music cognition [5]. These representational choices define the event vocabulary and its granularity, and therefore shape which regularities can be expressed or detected, including uncertainty about what comes next. Accordingly, what matters is not only the overall level of this uncertainty, but also how it is organized in a piece [2,3]. Since both depend on the chosen event vocabulary, representational choices can systematically reshape the uncertainty patterns we infer.

For example, consider a short melody of eight notes: A4–C5–A5–F5–A6–C4–F4–A7. If we encode each note as a distinct event by combining pitch class and octave, the sequence contains eight different event types, so each event type tends to have only one observed next-event option, making transitions appear highly predictable. If, instead, we collapse octave and keep only pitch class, the same melody reduces to the pitch-class sequence (A–C–A–F–A–C–F–A) and to just three event types (A, C, and F) repeated across the eight notes. Each event type can then be followed by different events (e.g., A can be followed by either C or F), increasing uncertainty. Thus, although both encodings describe the same musical surface, they induce different transition statistics, raising a basic methodological question: how does representational detail shape the structure we infer from music?

One way to make such differences explicit is to use network representations, in which sequence elements are

*Contact author: Lluc.Bono.Rossello@ulb.be

modeled as nodes and observed transitions between elements as directed edges. Previous work used pitch-based representations with octave and duration [6,7], pitch-class sets [8–12], and chord-split representations [13,14], which resulted in different topologies and statistical insights [15–19]. Yet, without a systematic comparison across levels of detail within the same musical corpus, it remains unclear whether the reported topological patterns reflect the music itself or the encoding, motivating a cross-representation analysis that makes explicit how different event vocabularies redistribute uncertainty across the network.

However, because expectation and surprise are perceptual phenomena, the relevant question is not only what uncertainty is present in the empirical transition structure, but also how reliably that structure can be internalized under imperfect memory and noise. Musical expectation is shaped by exposure: listeners learn statistical regularities from musical experience, and these learned regularities influence what later sounds expected or surprising [20,21]. If such transition statistics are not internalized perfectly, the structure available for prediction may be a biased approximation of the empirical sequence, and later expectations would inherit this distortion.

This raises a second question: under what conditions does a representation’s transition structure remain stable when it is internalized under perceptual constraints, and therefore provide reliable statistics for later expectation formation? Human responses to sequential input are limited by memory, noise, and representational cost [22–26]. A key consequence is that observers may not rely only on the immediately observed one-step transitions, but may partially integrate information over multiple steps—treating events that are reachable in two or more steps as if they were more directly available than the true statistics imply. Such integration can cause listeners to treat events that occur only after several steps as if they were possible next events, thereby changing the transition probabilities they infer. Recent work shows that the recoverability of transition statistics under such perceptual constraints depends on the organization of uncertainty [22]. Lynn *et al.* [23] showed, using reaction-time experiments with visual sequences, that even when two sequences have the same average uncertainty, meaning that continuations are equally unpredictable, differences in the organization of transitions can cause humans to react as if the uncertainty were higher, resulting in a perceptual error, defined as the mismatch between the true transition statistics and the observer’s inferred model.

Returning to our melodic example, this suggests that the pitch-class + octave encoding should allow for a single empirical continuation at each step, yet a learner with imperfect access to the sequence may still fail to recover this structure and treat other continuations as plausible. For instance, after observing an event such as A6, the true continuation in the sequence is C4. However, a biased estimate may also assign probability to an alternative, such as F4, consistent with the multistep integration described above, in which events reachable after multiple steps are treated as if they were directly available next. Introducing such competing continuations increases the uncertainty of the internalized transition structure and alters the inferred transition structure from a single-outgoing edge pattern to a branching one. Conversely, in the pitch-class-only representation, the same context re-

duces to A, for which both continuations C and F are in fact observed in the melody (i.e., A is followed by C and also by F). In that case, an error that mixes up which continuation occurred does not introduce a qualitatively new transition, but mainly shifts probability mass between already-available options (e.g., between $A \rightarrow C$ and $A \rightarrow F$), making the representation more robust to such distortion during internalization. Therefore, understanding how topology modulates this distortion of transition statistics is necessary for evaluating which representations provide transition structures that are more stable under perceptual constraints and may therefore serve as more reliable inputs for later prediction.

In short, the choice of representation fixes the transition statistics and therefore the distribution of uncertainty. Representing these statistics as a network makes their topology explicit. Because humans do not access transition statistics perfectly, perceptual constraints can make the same statistical uncertainty appear higher or lower, depending on how transitions are organized. This motivates (1) a systematic comparison of network structure across levels of representation detail and (2) an analysis of how the resulting network topology relates to distortion under perceptual constraints.

In this work, we address this issue by analyzing eight network representations of music, drawing on and extending existing approaches in the literature. These vary in their level of detail, from single-feature encodings (pitch, interval, duration) to multifeature combinations that incorporate octave or duration with pitch. We quantify how these representational choices reshape network topology and evaluate how efficiently each representation structures information by computing both the resulting entropy rate and the divergence from a perceptual-constraint estimate. Together, these measures make explicit a trade-off: increasing representational detail can preserve finer aspects of musical structure, but may produce a transition structure that is harder for humans to learn and use, while simpler encodings may sacrifice detail yet yield transition estimates that remain closer to the empirical statistics [22,27].

In addition, by further examining how uncertainty and modeled error are distributed locally across each network representation—and how these patterns vary with piece length—we identify conditions under which random-walk centrality aligns with uncertainty. By quantifying how topology drives divergences between true transition statistics and a perceptually constrained estimate, our approach both reveals a representation-dependent bias that ideal-access expectation models can overlook [21,28] and points toward incorporating realistic memory limits when relating uncertainty to neural [3] and behavioral data [4,29].

II. RESULTS

To examine how different levels of musical description shape structural and informational properties, we constructed eight network representations of the same sequences, spanning both established encodings from the literature and controlled feature-level variants introduced for systematic comparison. We start with minimal, single-feature encodings that expose the basic *building blocks* of musical structure, and then introduce additional features stepwise, allowing us to

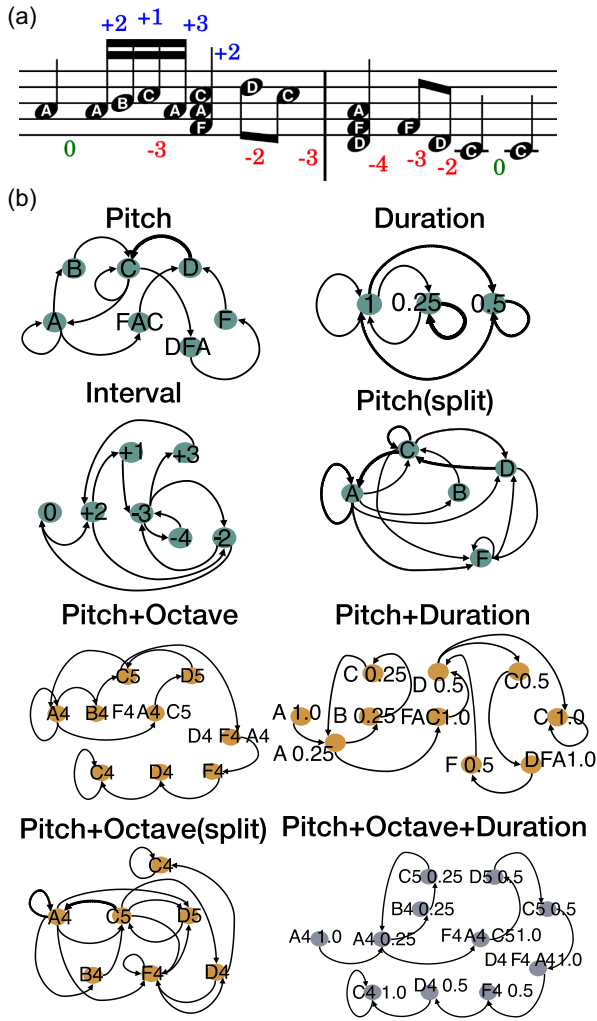


FIG. 1. Networks reconstructed from a musical piece. (a) A simple composition with highlighted interval changes. (b) Eight distinct network representations. All networks are directed and weighted: edges indicate transitions and weights correspond to their frequencies.

observe how topological and informational patterns emerge or transform as representational richness increases. In all cases, nodes correspond to unique feature configurations and edges record empirical transitions between consecutive elements in the musical sequence. The eight representations are defined as follows:

- (1) *Pitch*: Each node represents the note being played, identified by its pitch class (e.g., C, D, E), without distinguishing the octave. If several notes occur together (a chord), they form a single node [9].
- (2) *Duration*: Each node represents how long a note or chord lasts, expressed as a standardized value relative to a beat.
- (3) *Interval*: Each node represents the distance in semi-tones between two consecutive notes or chords [see Fig. 1(a)].
- (4) *Pitch+Duration*: Each node represents the note or chord together with its duration.
- (5) *Pitch+Octave*: Each node represents the note or chord together with its octave information.

(6) *Pitch+Duration+Octave*: Each node represents the note or chord with its duration and octave, combining all three features [6,7].

(7) *Pitch (split)*: A variation of the *Pitch* model in which chords are split, so that each note is represented as a separate node. Edges are created from the predecessor to each note in the chord and from each note to the possible successors.

(8) *Pitch+Octave (split)*: A variation of the *Pitch+Octave* model in which chords are split, with each note and its octave treated as a separate node. As in the previous model, edges connect the predecessor to each note in the chord and from each note to the possible successors [13,14].

Given a musical piece in MIDI format, we constructed networks by tracking transitions between consecutive musical elements. Each transition forms a directed edge, while the definition of a node depends on the specific model chosen. The resulting networks are both *directed* and *weighted*, with edge weights corresponding to transition counts during the execution of the piece. Figure 1 illustrates the differences between the different representations. The Supplemental Material [30] further illustrates these differences through network representations of Fryderyk Chopin's Waltz in E-flat major, Op. 18.

In this work, we focus on piano compositions, which are both musically rich across multiple dimensions and well documented in large, curated corpora. Our corpus combines two sources: the piano-midi.de dataset [31], which contains 268 pieces, and the MSMD dataset [32], which contains 665 pieces. Together, these datasets cover works ranging from the Baroque period to the 20th century. We preprocessed and merged the two collections, separating the musical scores by left and right hands. Further details are provided in Sec. IV A.

A. Structural consequences of representational choice

To assess how representational choices affect network topology, we first focused on four measures that summarize the main structural changes induced by the different event vocabularies: network size, average degree, degree heterogeneity, and clustering coefficient. Network size quantifies the number of distinct musical states generated by a representation. The average degree measures the average number of transitions per state, degree heterogeneity captures variability in the degree distribution, and the clustering coefficient measures the extent to which nodes in the graph tend to form tightly connected groups. These measures provide the structural baseline for the analysis since each representation defines a different set of musical states, and therefore a different transition network. Figure 2 shows these measures for the right-hand networks, while the Supplemental Material [30] reports additional topological measures and the corresponding left-hand results.

The clearest effect of increasing representational detail is an expansion of the state space. The most compressed encodings, such as *Duration* and *Pitch (split)*, produce the smallest networks, whereas combined encodings such as *Pitch+Duration* and *Pitch+Octave+Duration* produce the largest networks [Fig. 2(a)]. For most nonsplit representations, this expansion is accompanied by a decrease in average degree and clustering coefficient [Figs. 2(b) and 2(d)], indicating that

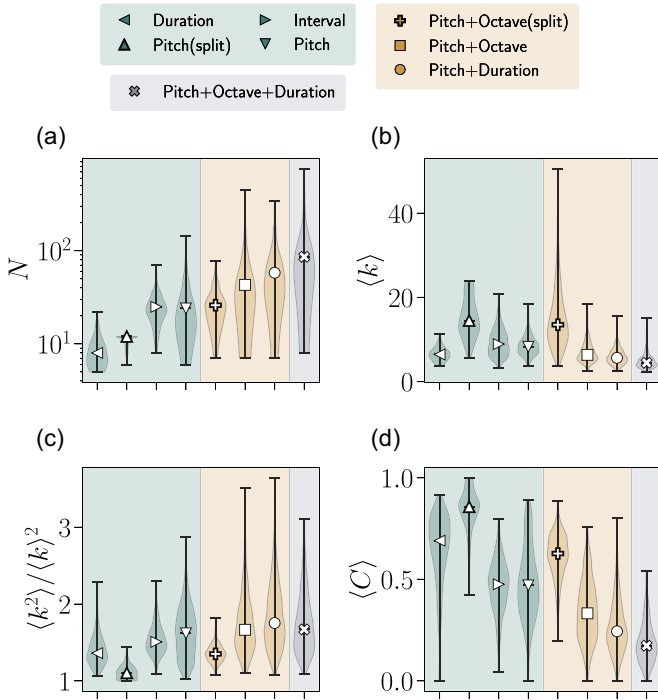


FIG. 2. Topological properties of music network models. Violin plots showing the distribution of four network measures: (a) size N , (b) average degree $\langle k \rangle$, (c) degree heterogeneity $\langle k^2 \rangle / \langle k \rangle^2$, and (d) average clustering coefficient $\langle C \rangle$, across different models averaged over all musical pieces for the right hand. Models are grouped by the number of musical features used. Each symbol corresponds to an individual model and also highlights the median of each distribution.

transitions become distributed across a larger and sparser set of states rather than concentrated among a small number of recurring event types.

The *Duration* representation is a limiting case: despite its highly compressed vocabulary, it does not become dense, because the observed duration-to-duration transitions occupy only a restricted subset of the possible state space. The split representations also depart from the general trend. By decomposing simultaneous events into separate note transitions, they create multiple local connections, which keeps them relatively dense and highly clustered even though their state spaces are not among the largest.

At the same time, the expansion is not a simple rescaling of the network. Degree heterogeneity increases as additional musical features are incorporated [Fig. 2(c)], showing that richer representations create a stronger contrast between highly connected states and more peripheral states. In musical terms, some events become transition hubs that participate in many possible continuations, whereas others occur in more specific sequential contexts. Split representations again deviate from this pattern: their local branching increases density and clustering, but it also produces more homogeneous degree distributions than the corresponding nonsplit representations. Thus, the split models should be interpreted as a distinct construction rather than as a simple increase in representational detail.

Together, these results show that representational choice changes more than the number of nodes in the network. Com-

pressed representations concentrate transitions into smaller and denser structures, whereas richer representations expand the state space into sparser and more heterogeneous networks. This structural contrast motivates the information-theoretic analyses below, where we move from the topology of each representation to the transition statistics that these topologies support.

B. Informational and perceptual effects of representational choice

Next, we ask what each representation makes available for prediction. Once a musical sequence is encoded as a transition network, each state carries information about what events can follow it. However, representational choice affects prediction along two dimensions: first, how strongly the current state constrains the next event; and second, according to recent work [22,23], how robust the resulting transition structure is when approximated under imperfect access to the sequence. We therefore analyze the transition networks in two steps. First, we quantify the entropy rate of each empirical transition structure. This measures, on average, how many continuations remain plausible when the sequence is followed from one state to the next. High entropy means that the current encoded state leaves several continuations plausible and is therefore less selective about what can follow; low entropy means that the current state more strongly constrains the next event. Thus, entropy is interpreted here as next-state uncertainty: a measure of how predictable the transition structure becomes under a given encoding. Second, we ask how this same structure changes when it is approximated through a biased transition estimate. This captures how empirical transition statistics may be distorted during internalization, before they can support later prediction. We model this distortion using a recently proposed framework for how transition probabilities may be internally approximated during abstract sequence learning [23].

First, following prior work on information flow in musical networks [13], we quantify the next-state uncertainty using the Shannon entropy rate of a random walk on the network [33,34] (see Sec. IV B),

$$S = \sum_i \pi_i S_i = - \sum_i \pi_i \sum_j P_{ij} \log_2 P_{ij}, \quad (1)$$

where P_{ij} is the probability of moving from state i to state j , S_i is the local entropy of the outgoing transitions from state i , and π_i is the stationary probability of visiting state i in the network. The stationary distribution is obtained from the transition matrix $\mathbf{P}$ and reflects how often a diffusive random walker encounters each node in the long run. The entropy rate, therefore, captures how next-state uncertainty is dynamically organized across the representation: it is high when frequently visited states have many plausible continuations, and low when frequently visited states have more constrained outgoing transitions. In information-theoretic terms, this quantity can be read as the average information produced per step of the encoded sequence; in the present analysis, we use this same quantity to ask how strongly the current encoded state constrains what can follow.

Second, we quantify how this empirical transition structure is altered by the biased transition estimate. Using the biased transition framework introduced by Lynn *et al.* [22,23], we

compare the empirical transition matrix $\mathbf{P}$ with a biased estimate $\hat{\mathbf{P}}$. The model produces this inferred transition matrix from the empirical transition matrix as

$$\hat{\mathbf{P}} = (\mathbf{I} - \eta)\mathbf{P}(\mathbf{I} - \eta\mathbf{P})^{-1}, \quad (2)$$

where $\eta \in [0, 1]$ controls the strength of the bias. Lower values of η keep the estimate closer to the direct one-step transition matrix, whereas higher values increase the contribution of indirect paths. This follows from the expansion

$$(\mathbf{I} - \eta\mathbf{P})^{-1} = \sum_{t=0}^{\infty} (\eta\mathbf{P})^t, \quad (3)$$

which shows that $\hat{\mathbf{P}}$ integrates not only direct transitions, but also higher-order paths of length 2, 3, or more, each weighted by a decaying factor. Consequently, $\hat{\mathbf{P}}$ can be interpreted as a temporally smoothed, biased estimate of the transition structure that would be internalized under imperfect access to the empirical transition statistics. In our analyses, we set ($\eta = 0.8$), following the empirical value determined by Lynn *et al.* [22].

We quantify the difference between the empirical and biased transition structures using the Kullback-Leibler (KL) divergence

$$D_{\text{KL}}(\mathbf{P} \parallel \hat{\mathbf{P}}) = - \sum_i \pi_i \sum_j P_{ij} \log_2 \frac{\hat{P}_{ij}}{P_{ij}}. \quad (4)$$

Lower values indicate that the biased estimate remains close to the empirical transition probabilities, whereas higher values indicate that the imposed bias distorts the representation's transition structure more strongly. In this sense, the model targets an early step in statistical learning: before transition statistics can support later predictions, they must first be internalized from experience, and this internalization may itself be biased.

Figure 3(a) shows that richer musical representations provide a more precise description of what can come next, but this precision makes their transition structure more vulnerable to distortion under the biased estimate. Across most representations, increasing the number of encoded musical features lowers next-state uncertainty while increasing internalization error, quantified by D_{KL} . Compressed representations such as *Pitch*, *Duration*, and *Pitch (split)* tend to occupy the low-divergence region: their empirical transition structures remain relatively close to the biased transition estimate. However, these representations also encode musical events coarsely, and in several cases retain relatively high entropy because many different continuations are collapsed onto the same state space. In contrast, richer representations such as *Pitch+Duration* and *Pitch+Octave+Duration* preserve finer distinctions between events. This added specificity tends to lower entropy by making transitions more constrained, but it also increases D_{KL} , indicating that the resulting transition structure is more strongly reshaped by the biased estimate.

The split representations should be interpreted separately from this ordering. By decomposing simultaneous events into separate note transitions, they inflate local branching and introduce additional short-range paths that are not present in the corresponding nonsplit representations. This effect is most

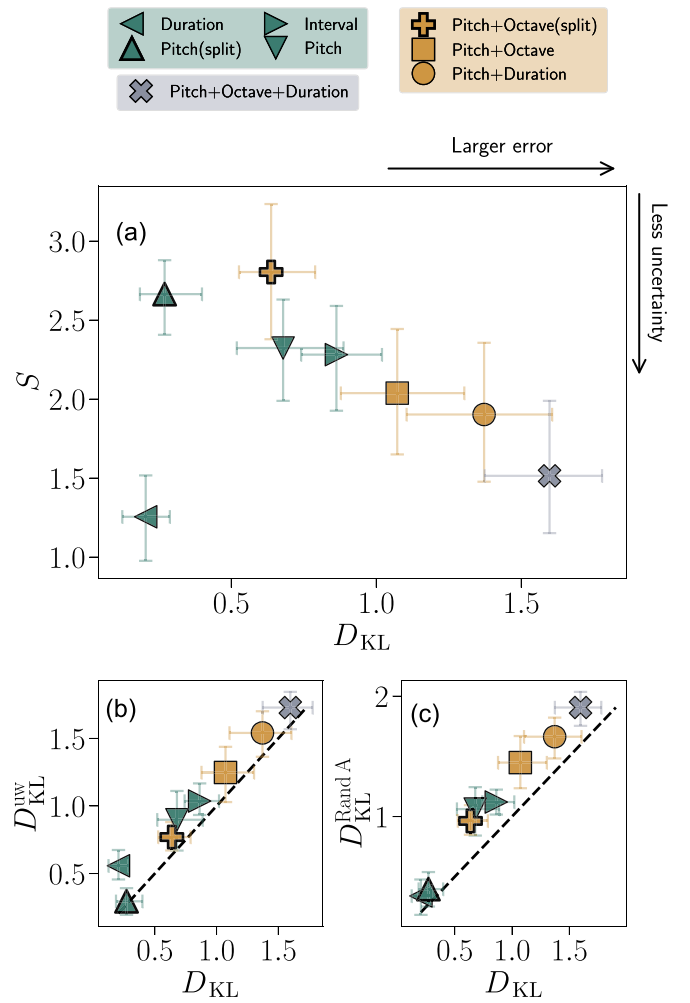


FIG. 3. Transition uncertainty and robustness to biased internalization across music network representations. (a) Median D_{KL} vs median S . (b) Median Kullback-Leibler divergence of weighted D_{KL} vs unweighted $D_{\text{KL}}^{\text{uw}}$ networks. (c) Median D_{KL} vs D_{KL} of type-A randomized networks. Models are grouped by the number of musical features used. Each symbol corresponds to an individual model. Error bars show the interquartile range around the median. All panels correspond to right-hand tracks.

compressed in *Pitch (split)*, where the state space is bounded by the 12 pitch classes. As a result, the network can become highly connected within a very small vocabulary, yielding high next-state uncertainty while leaving relatively little detailed transition structure for the biased estimate to disrupt. *Pitch+Octave (split)* shows the same construction effect in a larger state space: its position in Fig. 3(a) therefore reflects not only increased representational detail, but also the specific way in which chords are expanded into sequential transitions.

This pattern reveals a representational trade-off: richer vocabularies make continuations more specific, but their sharper transition profiles are more susceptible to distortion under biased estimation. Compressed vocabularies are more robust partly because they collapse distinctions that may be musically meaningful. The *Duration* representation is a limiting case, where low entropy and low divergence mainly reflect a very restricted state space.

The bottom panels of Fig. 3 test two aspects of this interpretation. First, Fig. 3(b) compares the weighted networks used throughout the main analysis with an unweighted version of the same network skeletons. In the weighted networks, transition probabilities reflect how often each transition occurs in the musical sequence. In the unweighted version, only the presence or absence of a transition is retained, so all outgoing transitions from a state are treated equally. This comparison, therefore, asks whether transition frequencies help preserve the empirical transition structure under the biased estimate, beyond the topology of possible transitions alone. The unweighted networks show systematically higher D_{KL} than the weighted networks, indicating that empirical transition frequencies reduce the distortion introduced by the biased estimate. Repeated transitions, therefore, do not merely add regularity in a trivial sense; they help the biased estimate remain closer to the original empirical transition structure. The same pattern is observed for the left hand, as reported in the Supplemental Material [30].

Second, Fig. 3(c) asks whether the divergence observed in the real musical networks is specific to their empirical organization. We compute D_{KL} for each real network by comparing its empirical transition matrix with its biased estimate, and repeat the same procedure for degree- and out-strength-preserving randomized counterparts (type-A randomization). These null models preserve local node-level constraints but disrupt the empirical arrangement of transitions across the network. The comparison tests whether divergence depends only on vocabulary size and local connectivity, or also on the global organization of the musical transition structure. Real networks show lower D_{KL} than their randomized counterparts, indicating that the empirical placement of transition weights and stationary visitation probability makes the transition structure less vulnerable to distortion by the biased estimate. Additional null-model analyses, including randomizations of edge weights on the same network skeleton (type-B randomization), are reported in the Supplemental Material [30].

Together, these results show that representational choice affects both the uncertainty of the empirical transition structure and its robustness to biased internalization. Thus, the trade-off is between simplified transition structures that are robust but less specific, and more precise transition structures that are richer as sequential descriptions but more costly to internalize. Event representation should therefore not be treated as a neutral preprocessing step: it determines which transition statistics are available for later learning or prediction, and how robust those statistics are during internalization.

C. Stationary flow separates uncertainty and divergence across visited states

The trade-off between entropy and KL divergence is a global property of each representation. However, both quantities are obtained by averaging local node-level values with the stationary distribution [see Eqs. (1) and (10)]. The stationary probability of a node indicates how often that state is visited in the long-run dynamics of the network, and therefore how strongly it contributes to the global measures. This means that the same set of local transitions can produce different

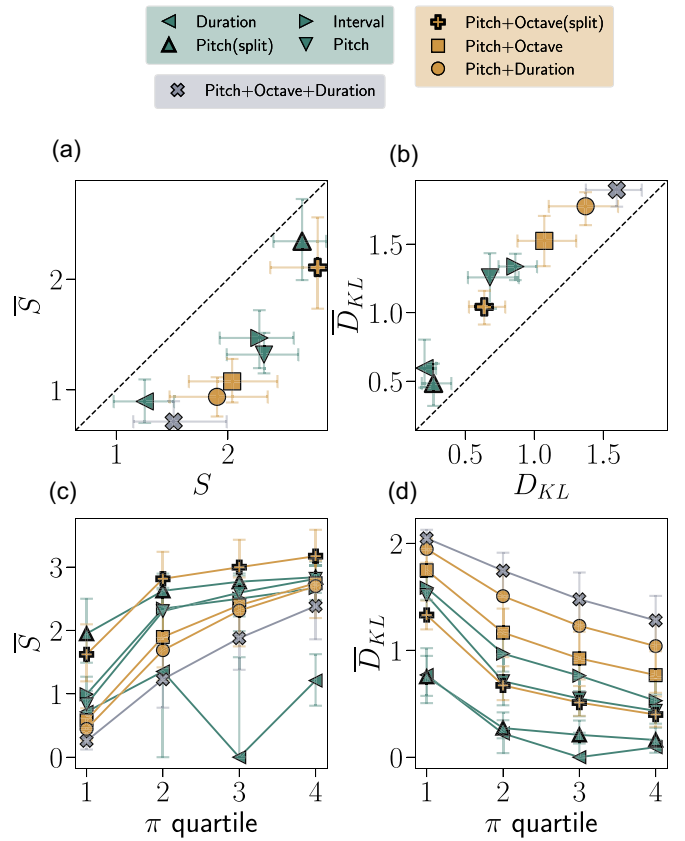


FIG. 4. Global and local alignment of uncertainty and inference in music networks. (a) Median entropy S vs mean node-level entropy $\bar{S}$. (b) Median KL divergence D_{KL} vs node-averaged KL divergence $\bar{D}_{KL}$. (c) Average node entropy ($\bar{S}$) across π quartiles. (d) Node-averaged KL divergence ($\bar{D}_{KL}$) across π quartiles. Points show medians and error bars indicate the interquartile range. All panels correspond to right-hand tracks. Models are grouped by the number of musical features used. Each symbol corresponds to an individual model.

global values depending on whether uncertainty or divergence is associated with high- or low-visitation states. We therefore next ask how local uncertainty and local divergence are distributed across the stationary visitation profile of each musical network.

To do so, we compared the entropy rate S computed above with the uniform node average $\bar{S} = \frac{1}{N} \sum_i S_i$. The uniform average treats all states equally, whereas S places greater weight on states visited more often by the random walk. Across representations, S is consistently larger than $\bar{S}$ [Fig. 4(a)]. This shows that high-visitation states tend to have higher local transition uncertainty than the average state. The global entropy rate is therefore determined not only by how uncertain local transitions are, but also by where this uncertainty is located along the stationary distribution.

We then applied the same comparison to the KL divergence. D_{KL} is consistently lower than the uniform node average $\bar{D}_{KL}$ [Fig. 4(b)]. Thus, states that contribute most strongly to the long-run dynamics tend to be those whose transition probabilities are less distorted by the biased transition estimate. Conversely, states with larger divergence tend

to have a lower stationary visitation and therefore contribute less to the global divergence. This also helps explain why empirical transition weights reduce D_{KL} in Fig. 3(b): weights not only define local transition probabilities, but also shape the stationary distribution, increasing the contribution of states whose transition structure is more robust to the bias.

Together, these two comparisons show that uncertainty and divergence are distributed asymmetrically across the stationary distribution of musical networks. Frequently visited states tend to carry more uncertainty but lower divergence under the modeled perceptual constraints, whereas less frequently visited states tend to carry less dynamical weight and higher divergence. From the perspective of biased internalization, this means that the transitions encountered most often are also those whose empirical probabilities are more closely preserved by the biased estimate.

A more local view helps to specify how this organization varies across the stationary distribution. We grouped the nodes into quartiles according to their stationary visitation probability and examined local entropy and local KL divergence within each quartile. The highest-visitation quartiles carry the highest local uncertainty and the lowest local divergence, whereas the lowest-visitation quartiles tend to show the opposite pattern [Figs. 4(c) and 4(d)]. Because the stationary distribution is itself highly skewed, global entropy and global divergence are shaped primarily by the subset of states most frequently visited by the random walk rather than by the full state space uniformly. The Supplemental Material [30] reports the corresponding stationary-distribution analyses and left-hand results.

D. Piece length strengthens visitation-based organization of uncertainty and divergence

Having shown that uncertainty and divergence vary across stationary visitation levels, we next ask whether this visitation-based organization depends on the amount of musical material available in a piece. Longer pieces contain more transitions and may therefore provide more opportunities for frequently visited and rarely visited states to acquire different informational roles.

To quantify this effect, we defined two visitation-contrast measures by comparing each stationary-weighted quantity with its corresponding uniform node average. For entropy, we used the entropy visitation contrast,

$$C_S = S - \bar{S}, \quad (5)$$

which is positive when frequently visited states have higher local entropy than the average state. For KL divergence, we used the divergence visitation contrast,

$$C_{KL} = D_{KL} - \bar{D}_{KL}, \quad (6)$$

which is negative when frequently visited states have lower KL divergence than the average state. Thus, both quantities are defined as stationary-weighted minus uniform averages, but their expected signs reflect the asymmetric organization observed above: uncertainty is enriched in frequently visited states, whereas divergence is reduced in frequently visited states.

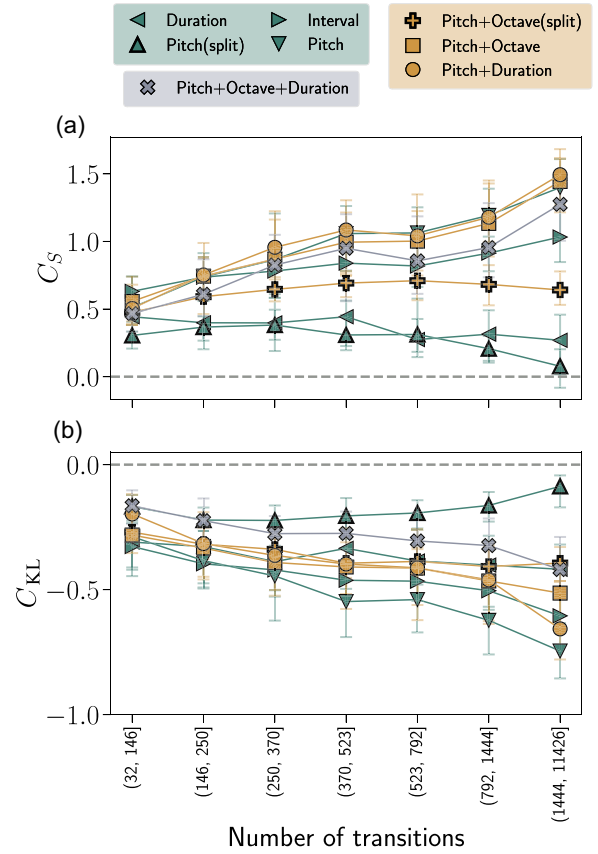


FIG. 5. Impact of piece length on the organization of uncertainty and inference in music networks. (a) Entropy-based visitation contrast C_S as a function of transition count. (b) Divergence-based visitation contrast C_{KL} as a function of transition count. Transition counts are divided into seven equal-sized bins. Points indicate medians and error bars show the interquartile range. All panels correspond to right-hand tracks.

As the number of transitions in a piece increases, the two visitation-contrast measures move in opposite directions (Fig. 5). The entropy visitation contrast C_S becomes more positive, indicating that longer pieces increasingly concentrate stationary visitation on states with richer local branching [Fig. 5(a)]. In contrast, the divergence visitation contrast C_{KL} becomes more negative, indicating that longer pieces increasingly concentrate visitation on states whose empirical transition probabilities remain closer to the biased transition estimate [Fig. 5(b)]. Thus, the effect of length is not only to add more transitions or enlarge the state space. Rather, with more musical material, the network develops a stronger visitation-based separation between where uncertainty is carried and where distortion under the bias is concentrated. The corresponding left-hand analysis is reported in the Supplemental Material [30].

This length dependence differs across representations. In highly compressed or split representations, the trends are weaker and less consistent: the state space is either too small or too locally constrained for a strong visitation-based separation to emerge. In richer multifeature representations, in contrast, increasing piece length produces a clearer increase in C_S and a clearer decrease in C_{KL} . This indicates that the

organization observed in Fig. 5 becomes stronger when the representation contains both sufficient descriptive detail and enough transitions to support differentiated state roles. It also helps explain why these representations differ from the randomized networks discussed above: the empirical sequence does not merely populate a vocabulary, but organizes transition frequencies so that frequently visited and rarely visited states play different informational roles.

Overall, these results explain how the entropy-divergence trade-off develops within individual pieces. Representational detail determines the available state space, while piece length shapes how strongly uncertainty and divergence are separated across the stationary distribution. Thus, longer pieces produce more structured informational landscapes, in which frequently visited states carry greater uncertainty while remaining comparatively less distorted under the biased transition estimate.

III. DISCUSSION

Representational choice plays a decisive role in the structural properties of musical networks. Our results show that the number and grouping of features—and whether musical events are split or not—substantially influence the resulting topology. Properties often associated with musical corpora, such as small-world organization [6,7], heavy-tailed degree distributions, or Zipf- and Heaps-like scaling [11,16], emerge only under certain representational levels and are not inherent to the raw data alone. More detailed, multifeature representations tend to produce markedly heterogeneous and skewed structures, whereas compressed vocabularies attenuate these patterns. These findings show that topological regularities in music depend on the definition of the underlying symbolic space.

Building on this, our results frame representational differences as a trade-off between descriptive richness and robustness to distortion under perceptual constraints. Here, descriptive richness denotes the number of musically relevant distinctions preserved by an encoding. This determines how large and how fine-grained its event vocabulary, or state space, is. A richer encoding typically makes one-step transitions more specific and more locally predictable; i.e., it lowers next-state uncertainty. However, it can also make the transition structure more vulnerable to distortion when transition statistics are estimated under perceptual-constraint models [23]. Robustness to perceptual distortion refers to how closely the constrained transition estimate remains to the empirical transition structure. This captures a possible bias in the internalization of transition statistics, rather than the full process by which listeners form predictions. Compressed representations tend to yield lower distortion, suggesting that their transition topology is less altered by the constrained estimate. In contrast, richer representations that preserve finer details reshape the state space and introduce more specific transition structures that are harder to estimate, increasing the mismatch between empirical and constrained transition estimates. This framing also clarifies how our interpretation relates to previous work on information processing in complex and musical networks. Lynn *et al.* [22] and Kulkarni *et al.* [13] use entropy rate and divergence to characterize how efficiently transition structures communicate or process information. We

use the same formal quantities but emphasize their meaning from the perspectives of prediction and internalization. A high entropy rate means that the current encoded state leaves many continuations possible. In information-theoretic terms, this corresponds to high information per step, but predictively it also implies that the state is less selective: knowing the current state does not strongly constrain what can come next. Such diffuse transition structures can be easier for the biased estimate to approximate, because many alternatives are already present in the empirical transition matrix. Conversely, a lower entropy rate can arise when a richer representation defines more specific states with more constrained continuations. These states are more useful for prediction because they say more about what is likely to follow, but their sharper transition profiles are also more vulnerable to distortion when indirect paths are mixed into the biased estimate. The increase in D_{KL} for richer encodings should therefore not be interpreted as a cost of reduced information, but as the cost of internalizing a more specific and more predictive transition architecture. This echoes principles from information bottleneck theory [27] and minimum-description-length formulations [35], in which representations balance fidelity against cognitive cost. In our network framework, this cost becomes explicit: the representation choice determines both the next-state uncertainty induced by the empirical transition matrix and how strongly that transition structure is distorted when the empirical matrix is approximated by a constrained estimate. Viewed this way, representational choices and perceptual constraints jointly determine which transition statistics may be available for later learning or prediction.

Although our analysis remains descriptive rather than cognitive, it helps specify a step that is often implicit in statistical-learning accounts of musical expectation: the acquisition of transition statistics from experience. Human expectations depend on learning and exposure [21,36], and therefore cannot be reduced to the empirical transitions of a single piece. Our results do not model cognition directly, but they show how the transition statistics available for later prediction may already be altered before prediction occurs. State-of-the-art models of musical expectation [28] estimate uncertainty from learned transitions, often assuming that the relevant transition statistics are available without distortion. In this work, we use network representations to make explicit how imperfect access to transition statistics may reshape the statistics available for learning, and to quantify how this effect depends on representational choice by measuring the topology-dependent gap between empirical and constrained transition estimates. A natural extension would be to incorporate this kind of biased transition internalization into statistical-learning models of musical expectation, and test whether predictions based on constrained transition estimates better explain behavioral or neural responses than predictions based on ideal access to empirical statistics. Such an extension would also make it possible to ask whether the visitation-based organization identified here relates to affective responses to music, where uncertainty, surprise, and predictability have been linked to engagement and pleasure [3,20,29].

Our representational hierarchy was chosen to build on established symbolic viewpoints in music-network research

while enabling a controlled progression from coarse to detailed descriptions. Other musically informed viewpoints—such as tonality-relative encodings based on scale degree or harmonic function (e.g., tonic/dominant), or learned feature spaces—may yield different structural signatures and should be explored in future work. In particular, future work could ask whether combining multiple musical dimensions within a single predictive framework reduces the trade-off observed here, or whether perceptual distortion persists when pitch, duration, harmony, and longer-range context are learned jointly. Similarly, our focus on classical piano music provided a rich context for examining how complexity grows with representational detail, but the same framework can be applied to other repertoires. Our intuition is that genres with more repetitive or minimalist structures will show weaker differentiation across representational levels, offering a useful contrast to the dynamics observed here.

More broadly, this approach demonstrates how the structural and informational consequences of representational choice can be analyzed in any sequential domain. Music serves as a clear test bed because its feature combinations and hierarchical abstractions are well characterized, yet the framework is readily transferable to other sequences and systems. By tracing how topologies and informational landscapes evolve with representational detail, we gain a principled way to compare how different feature sets shape the organization of sequential information and the robustness of learned transition statistics under constrained access.

IV. METHODS

A. Datasets and preprocessing

We used two publicly available collections of piano music in MIDI format: the piano-midi.de dataset [31] (268 pieces) and the MSMD dataset [32] (665 pieces). After merging, the corpus comprised 933 files spanning works from the Baroque to the 20th century. Some duplicates remain, since different encodings of the same composition may capture distinct transcription choices.

MIDI files were selected only if they included tempo and time-signature metadata to ensure accurate quantization of durations. We further restricted the corpus to files containing exactly two tracks, corresponding to the left and right hands of a piano score. Each track was converted into a directed and weighted network, resulting in two networks for each composition. To ensure a well-defined entropy with a unique stationary distribution, we restrict our analysis to the largest strongly connected component (LSCC) of each network. Moreover, to ensure that the resulting networks faithfully represented the musical pieces, we retained only LSCCs that contained at least 80% of the nodes, yielding more than 600 networks per model.

During preprocessing, each musical element was encoded according to the chosen network representation (see Sec. II). Symbolic features include pitch (frequency, grouped into pitch classes an octave apart), duration (quantized relative to a quarter note = 1.0), and chromatic interval (pitch distance in semitones between consecutive events). If multiple pitches occurred simultaneously, the highest pitch was used to compute

intervals. Descriptive statistics on pitch ranges, note durations, chord distributions, and interval patterns for both datasets, including systematic differences between left and right hands, are provided in the Supplemental Material [30].

B. Entropy in complex networks

The Shannon entropy for a node i can be written as [13,34]. All logarithms are taken to base 2, so entropy and KL divergence are reported in bits.

$$S_i = - \sum_j P_{ij} \log_2 P_{ij}, \quad (7)$$

where P_{ij} are the entries of the transition matrix $\mathbf{P}$. In the case of directed and weighted networks, $P_{ij} = \omega_{ij}/s_i^{\text{out}}$, where ω_{ij} is the edge weight and s_i^{out} is the out-strength of the node i . Here, we consider only outgoing links from node i . In unweighted networks, the out-strength is replaced by the out-degree k_i^{out} .

The entropy of the entire network has the form of Eq. (1). The stationary distribution π satisfies the condition $\mathbf{P}\pi = \pi$ and can be obtained as the normalized eigenvector of $\mathbf{P}$ corresponding to the eigenvalue 1. For the simplest case, i.e., an undirected and unweighted graph, the stationary distribution becomes $\pi_i = k_i/2E$, where k_i is the degree of node i and E is the total number of edges. It is worth noting that the formula is only applicable when a network has a single strongly connected component. In the preprocessing step, we ensure that we work with such networks.

C. Human perception model

We adopt the human perception model introduced by Lynn *et al.* [23], which we briefly summarize here for completeness, as it serves as the basis for our perceptual analysis. In this model, the inferred transition probabilities $\hat{\mathbf{P}}$ are expressed in terms of the true transition probabilities $\mathbf{P}$ as

$$\hat{\mathbf{P}} = (\mathbf{I} - \eta)\mathbf{P}(\mathbf{I} - \eta\mathbf{P})^{-1}, \quad (8)$$

where $\eta \in [0, 1]$ controls the inaccuracy of perceptual expectations and $(\mathbf{I} - \eta\mathbf{P})^{-1}$ denotes a *matrix inverse*. This inverse can be expanded as a geometric series,

$$(\mathbf{I} - \eta\mathbf{P})^{-1} = \sum_{t=0}^{\infty} (\eta\mathbf{P})^t, \quad (9)$$

which reveals that perception integrates not only direct (one-step) transitions but also higher-order paths—two, three, or more steps apart—each weighted by a decaying factor η^t . Consequently, $\hat{\mathbf{P}}$ represents a temporally smoothed higher-order estimate of the transition structure. Empirical evidence suggests an average human accuracy parameter of $\eta \approx 0.8$, which we use for our computation [22].

To quantify the discrepancy in the perception of music networks, i.e., the difference between these two probabilities, we use the Kullback-Leibler divergence defined as

$$D_{\text{KL}}(\mathbf{P} \parallel \hat{\mathbf{P}}) = - \sum_i \pi_i \sum_j P_{ij} \log_2 \frac{\hat{P}_{ij}}{P_{ij}}. \quad (10)$$

The lower the KL divergence, the smaller the difference between $\mathbf{P}$ and $\hat{\mathbf{P}}$. In addition, we define the uniform average of the KL divergence as

$$\bar{D}_{\text{KL}}(\mathbf{P} \parallel \hat{\mathbf{P}}) = -\frac{1}{N} \sum_{i=1}^N \sum_j P_{ij} \log_2 \frac{\hat{P}_{ij}}{P_{ij}}. \quad (11)$$

ACKNOWLEDGMENTS

We thank Timoteo Carletti and Nicolás Bono Rosselló for helpful comments and feedback on the manuscript. L.B.R. and R.J. acknowledge support from the Bridge grant funded by Young Researchers of the Complex Systems Society (yrCSS). R.J. acknowledges support from the fellowship FI-SDUR funded by Generalitat de Catalunya. L.B.R. acknowledges support from the ARIAC project (No. 2010235), funded by the Service Public de Wal-

lonie (SPW Recherche). M.A.S. and M.B. acknowledge support from Grant No. PID2022-137505NB-C22 funded by MCIN/AEI/10.13039/501100011033 and by “ERDF/EU”. M.B. acknowledges the ICREA Academia award, funded by the Generalitat de Catalunya.

DATA AVAILABILITY

The data that support the findings of this article are openly available [31,32]. The code used to construct the network representations and reproduce the main analyses is available in Ref. [37]. The repository includes scripts for reading MIDI files from the analyzed datasets, extracting the viewpoint sequences used to define each representation, generating the corresponding directed weighted networks for each MIDI track, and reproducing the analyses reported in the main text.

- [1] S. Koelsch, Brain correlates of music-evoked emotions, *Nat. Rev. Neurosci.* **15**, 170 (2014).
- [2] S. Koelsch, P. Vuust, and K. Friston, Predictive processes and the peculiar case of music, *Trends Cognit. Sci.* **23**, 63 (2019).
- [3] V. K. M. Cheung, P. M. C. Harrison, L. Meyer, M. T. Pearce, J.-D. Haynes, and S. Koelsch, Uncertainty and surprise jointly predict musical pleasure and amygdala, hippocampus, and auditory cortex activity, *Curr. Biol.* **29**, 4084 (2019).
- [4] E. Mas-Herrero and J. Marco-Pallarés, Predictive processes shape individual musical preferences, *Proc. Natl. Acad. Sci. USA* **122**, e2500494122 (2025).
- [5] D. Conklin and I. H. Witten, Multiple viewpoint systems for music prediction, *J. New Music Res.* **24**, 51 (1995).
- [6] X. Liu, C. K. Tse, and M. Small, Composing music with complex networks, in *International Conference on Complex Sciences* (Springer, Berlin, Heidelberg, 2009), pp. 2196–2205.
- [7] S. Ferretti, On the complex network structure of musical pieces: Analysis of some use cases from different music genres, *Multimed. Tools Appl.* **77**, 16003 (2018).
- [8] F. Gomez, T. Lorimer, and R. Stoop, Complex networks of harmonic structure in classical music, in *International Conference on Nonlinear Dynamics of Electronic Systems* (Springer, Cham, Berlin, 2014), pp. 262–269.
- [9] J. Serrà, Á. Corral, M. Boguñá, M. Haro, and J. L. Arcos, Measuring the evolution of contemporary Western popular music, *Sci. Rep.* **2**, 521 (2012).
- [10] M. B. Nardelli, Tonal harmony and the topology of dynamical score networks, *J. Math. Music* **17**, 198 (2023).
- [11] B. Manaris, J. Romero, P. Machado, D. Krehbiel, T. Hirzel, W. Pharr, and R. B. Davis, Zipf’s law, music classification, and aesthetics, *Comput. Music J.* **29**, 55 (2005).
- [12] M. Serra-Peralta, J. Serrà, and Á. Corral, Heaps’ law and vocabulary richness in the history of classical music harmony, *EPJ Data Sci.* **10**, 40 (2021).
- [13] S. Kulkarni, S. U. David, C. W. Lynn, and D. S. Bassett, Information content of note transitions in the music of J. S. Bach, *Phys. Rev. Res.* **6**, 013136 (2024).
- [14] N. Di Marco, E. Loru, A. Galeazzi, M. Cinelli, and W. Quattrociochi, Decoding the evolution of melodic and harmonic structure of Western music through the lens of network science, *Sci. Rep.* **16**, 11121 (2026).
- [15] M. B. del Río, G. Cocho, and G. G. Naumis, Universality in the tail of musical note rank distribution, *Physica A* **387**, 5552 (2008).
- [16] D. H. Zanette, Zipf’s law and the creation of musical context, *Musicae Sci.* **10**, 3 (2006).
- [17] F. C. Moss, M. Neuwirth, D. Harasim, and M. Rohrmeier, Statistical characteristics of tonal harmony: A corpus study of Beethoven’s string quartets, *PLoS ONE* **14**, e0217242 (2019).
- [18] S. A. Mehr, M. Singh, D. Knox, D. M. Ketter, D. Pickens-Jones, S. Atwood, C. Lucas, N. Jacoby, A. A. Egner, E. J. Hopkins, *et al.*, Universality and diversity in human song, *Science* **366**, eaax0868 (2019).
- [19] M. Rohrmeier and I. Cross, Statistical properties of tonal harmony in Bach’s chorales, in *Proceedings of the 10th International Conference on Music Perception and Cognition* (ICMPC, Sapporo, 2008), Vol. 6, pp. 123–1319.
- [20] D. Huron, *Sweet Anticipation: Music and the Psychology of Expectation* (MIT Press, Cambridge, MA, 2008).
- [21] M. T. Pearce, Statistical learning and probabilistic prediction in music cognition: Mechanisms of stylistic enculturation, *Ann. NY Acad. Sci.* **1423**, 378 (2018).
- [22] C. W. Lynn, L. Papadopoulos, A. E. Kahn, and D. S. Bassett, Human information processing in complex networks, *Nat. Phys.* **16**, 965 (2020).
- [23] C. W. Lynn, A. E. Kahn, N. Nyema, and D. S. Bassett, Abstract representations of events arise from mental errors in learning and memory, *Nat. Commun.* **11**, 2313 (2020).
- [24] I. Momennejad, E. M. Russek, J. H. Cheong, M. M. Botvinick, N. D. Daw, and S. J. Gershman, The successor representation in human reinforcement learning, *Nat. Hum. Behav.* **1**, 680 (2017).
- [25] M. W. Howard and M. J. Kahana, A distributed representation of temporal context, *J. Math. Psychol.* **46**, 269 (2002).
- [26] F. Meyniel, M. Maheu, and S. Dehaene, Human inferences about sequences: A minimal transition probability model, *PLoS Comput. Biol.* **12**, e1005260 (2016).

- [27] N. Tishby, F. C. Pereira, and W. Bialek, The information bottleneck method, in *Proceedings of the 37th Annual Allerton Conference on Communication, Control, and Computing*, edited by B. Hajek and R. S. Sreenivas (University of Illinois, Urbana-Champaign, IL, 1999), pp. 368–377.
- [28] M. T. Pearce and G. A. Wiggins, Auditory expectation: The information dynamics of music perception and cognition, *Top. Cognit. Sci.* **4**, 625 (2012).
- [29] B. P. Gold, M. T. Pearce, E. Mas-Herrero, A. Dagher, and R. J. Zatorre, Predictability and uncertainty in the pleasure of music: A reward for learning? *J. Neurosci.* **39**, 9397 (2019).
- [30] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/x6hc-jb2r> for additional details on the network representations and supplementary figures.
- [31] B. Krueger, Piano MIDI, 1996–2018, <http://piano-midi.de/>.
- [32] M. Dorfer, H. Jan, Jr., A. Arzt, H. Frostel, and G. Widmer, MSMD—multimodal sheet music dataset, Zenodo, 2019, <https://doi.org/10.5281/zenodo.2597505>.
- [33] C. E. Shannon, A mathematical theory of communication, *Bell Syst. Tech. J.* **27**, 379 (1948).
- [34] J. Gómez-Gardenes and V. Latora, Entropy rate of diffusion processes on complex networks, *Phys. Rev. E* **78**, 065102(R) (2008).
- [35] J. Rissanen, A universal prior for integers and estimation by minimum description length, *Ann. Stat.* **11**, 416 (1983).
- [36] T. Daikoku, G. A. Wiggins, and Y. Nagai, Statistical properties of musical creativity: Roles of hierarchy and uncertainty in statistical learning, *Front. Neurosci.* **15**, 640412 (2021).
- [37] https://github.com/llucbono/network_representations.