

Information content of note transitions in the music of J. S. Bach

Suman Kulkarni ^{1,*} Sophia U. David,^{2,3} Christopher W. Lynn,^{4,5} and Dani S. Bassett^{1,2,6,7,8,9,†}

¹*Department of Physics & Astronomy, College of Arts & Sciences, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA*

²*Department of Bioengineering, School of Engineering & Applied Science, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA*

³*Department of Psychology, Yale University, New Haven, Connecticut 06520, USA*

⁴*Initiative for the Theoretical Sciences, Graduate Center, City University of New York, New York, New York 10016, USA*

⁵*Joseph Henry Laboratories of Physics, Princeton University, Princeton, New Jersey 08544, USA*

⁶*Department of Electrical & Systems Engineering, School of Engineering & Applied Science, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA*

⁷*Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA*

⁸*Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA*

⁹*Santa Fe Institute, Santa Fe, New Mexico 87501, USA*



(Received 6 September 2023; accepted 6 December 2023; published 2 February 2024)

Music has a complex structure that expresses emotion and conveys information. Humans process that information through imperfect cognitive instruments that produce a gestalt, smeared version of reality. How can we quantify the information contained in a piece of music? Further, what is the information inferred by a human, and how does that relate to (and differ from) the true structure of a piece? To tackle these questions quantitatively, we present a framework to study the information conveyed in a musical piece by constructing and analyzing networks formed by notes (nodes) and their transitions (edges). Using this framework, we analyze music composed by J. S. Bach through the lens of network science, information theory, and statistical physics. Regarded as one of the greatest composers in the Western music tradition, Bach's work is highly mathematically structured and spans a wide range of compositional forms, such as fugues and choral pieces. Conceptualizing each composition as a network of note transitions, we quantify the information contained in each piece and find that different kinds of compositions can be grouped together according to their information content and network structure. Moreover, using a model for how humans infer networks of information, we find that the music networks communicate large amounts of information while maintaining small deviations of the inferred network from the true network, suggesting that they are structured for efficient communication of information. We probe the network structures that enable this rapid and efficient communication of information—namely, high heterogeneity and strong clustering. Taken together, our findings shed light on the information and network properties of Bach's compositions. More generally, our simple framework serves as a stepping stone for exploring further musical complexities, creativity, and questions therein.

DOI: [10.1103/PhysRevResearch.6.013136](https://doi.org/10.1103/PhysRevResearch.6.013136)

I. INTRODUCTION

From Tibetan throat singing to Scottish piobaireachd to modern hip-hop, music is a universal aspect of human culture, enjoyed by people of all ages from all around the world. It has even been proposed that music is a fundamental part of being human [1]. Though styles, sounds, and instruments vary drastically from one culture and time period to another, it is indisputable that music has had a substantial impact on the development of humans and society [2,3]. Through music we can tell stories [4], convey messages [5], and imbue the

strongest of emotions [6–8]. It is a common human experience to feel pensive or despondent after hearing a slow song in a minor key or to feel carefree or energized after hearing an upbeat song in a major key. But how does something as abstract as music communicate so much? Past literature has discussed music in terms of expectation and surprise [9–13]. To be evolutionarily successful, our brains are adept at forming expectations based on prior events. When these expectations are contradicted by an experience, we feel surprised. With surprise can come a host of other emotions: We may feel relief when the dissonant sound we expected was actually consonant or we may feel distress when the musical resolution we expected did not occur [12–16]. But how do we quantify these expectations and surprises? How do we mathematically formalize and measure the information conveyed by a piece of music? Fundamentally, music is comprised of fleeting and elusive sounds, and hence may appear hard to measure.

Here, we seek to extract order from music's complexity by examining music through the lens of network science. A

*Corresponding author: sumank@sas.upenn.edu

†Corresponding author: dsb@seas.upenn.edu

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

network consists of nodes and edges—representing entities and the connections between them, respectively. Conceptualizing each note as a node and each transition between two notes as an edge, we can build a network for any piece of music [17–21]. This representation enables us to use physics-based approaches to quantitatively analyze aspects of a musical piece. Using music networks, we build a framework to study the information conveyed by a piece and apply this framework to provide a comprehensive analysis of Bach’s compositions. Bach is a natural case study given his prolific career, the wide appreciation his compositions have garnered, and the influence he had over contemporaneous and subsequent composers. His diverse compositions (from chorales to fugues) for a wide range of musicians (from singers to orchestra members) often share a fundamental underlying structure of repeated—and almost mathematical—musical themes and motifs. These features of Bach’s compositions make them particularly interesting to study using a mathematical framework.

As we listen to music, we form expectations [9,11,12,15]. Upon hearing a particular note, we anticipate which notes might come next based on past transitions. The less likely the outcome, the more surprised we are upon hearing it. This “surprisal” can be quantified by the Shannon information entropy [22]. Ideas from information theory have led to illuminating insights in a wide range of settings, including language [23,24], social networks [25,26], transportation patterns [27], and music [28,29]. We draw upon these ideas to quantify the information present in the music networks. While prior research has attempted to quantitatively identify patterns and features present across different kinds of music [19,30–32], understanding how humans perceive these patterns is more nuanced and complex than simply evaluating the structure of compositions because humans are not perfect learners. Rather, studies have consistently found that humans assimilate patterns of information presented to them through imperfect perceptual systems, resulting in slightly inaccurate representations of transition structures [33–36]. This observation raises interesting questions about the information that is perceived by a human; in particular, how does the inferred structure relate to, and differ from, the true structure of a musical piece? Further, are there any patterns in music that particularly shine through the messy process of human perception and, if so, how do these patterns vary across different kinds of music? While these questions are nuanced and can depend on factors like level of training [37–40], musical cultural conditioning [38,41–44], and even language [45–48], recent advances in the study of how humans learn networks of information offer a valuable framework to address these questions [34,35,49–51].

Here we draw upon ideas from network science, information theory, and cognitive science to build a framework to investigate the information conveyed by music. We then use this framework to provide a systematic analysis of music composed by J. S. Bach. We begin in Sec. II with a discussion of how music can be represented as a network along with details of the compositions analyzed in our work. Next, in Sec. III, we study the information present in the networks. We find that Bach’s music networks contain more information than expected from typical (or random) transition structures. Strikingly, we also find that certain composition forms are clustered together based on their information content. We

investigate how the network structure influences information content and show that the higher information in these music networks and the differences observed across musical pieces within each compositional form can be explained by the heterogeneity in node degrees (or the number of distinct pitches that follow a given note). Next, in Sec. V, we use a computational model for how humans learn networks of information to examine how closely the inferred transition structure of a piece aligns with the true network structure. We hypothesize that the music networks maintain a low deviation between the inferred and true network, and this property is driven by tight clustering in the network. Additionally, we find that certain compositional forms can be distinguished based on the discrepancies between the original and the inferred network. Together, our framework introduces a fresh perspective on music, and sheds light on properties of Bach’s music. By performing a systematic study of how information in a complex system, like music, is structured and perceived by humans, our paper provides insights on human creativity and how humans experience the world around them. Our study also opens up numerous interesting directions for further inquiry, which we outline in Sec. VII. Additionally, we emphasize the limitations of our analysis and discuss how future work could improve upon this paper to incorporate more realism in Sec. VII.

II. MUSIC AS A NETWORK OF NOTE TRANSITIONS

We note that there have been previous efforts in constructing and analyzing different network representations of music [17–21]. In our paper, we focus on investigating the information conveyed by note transitions in music and begin with a basic representation of the note transitions. We study a wide range of Bach’s compositions, including preludes, fugues, inventions, cantatas, English suites, French suites, chorales, Brandenburg concertos, toccatas, and concertos. The audio files for these pieces were collected and read in MIDI format, from which the sequence of notes was extracted (see Methods Sec. A 1 for further details on each compositional type and the sources for each piece). Each note present in a piece is represented as a node in the network, with notes from different octaves represented as distinct nodes. The transitions between notes are calculated separately for different instruments. If there is a transition from note i to note j , then we draw a *directed edge* from node i to node j (see Fig. 1). For chords, where multiple notes occur at the same time, edges are drawn between all notes in the first chord to all notes in the second chord. To simplify our analysis, we remove any self-loops in the network, thereby restricting ourselves to understanding the structure of transitions to the next *different* note in the piece. We begin by examining unweighted networks of note transitions to focus on how the network structure alone impacts the information content and perception of a musical piece. After understanding the skeleton of the transitions, we then add weights to the edges based on how frequently various transitions occur. This procedure allows us to disentangle the effects of the network structure (comprising the set of possible note transitions) and edge weights (comprising the note transition probabilities). Examples of note transition networks constructed using our simplified representation are illustrated in Fig. 2. Although our emphasis has been on building a

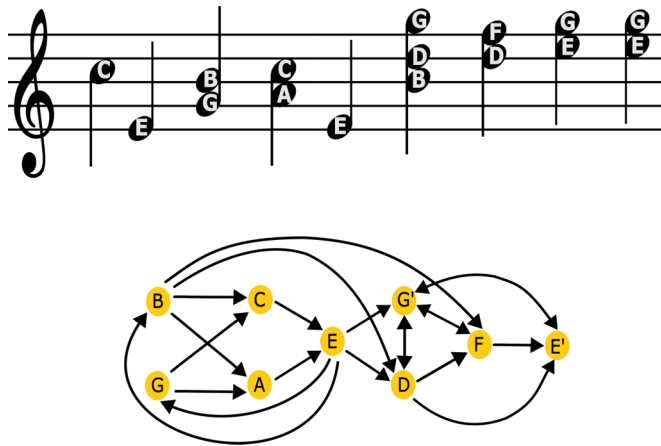


FIG. 1. An example of a network constructed from a musical piece using the method described in our paper. At the top, we show a toy musical piece. Below, we show the network in which notes are nodes and transitions between notes, whether isolated or played simultaneously as part of a chord, are directed edges. The direction of the edge matches the temporal direction of the transition.

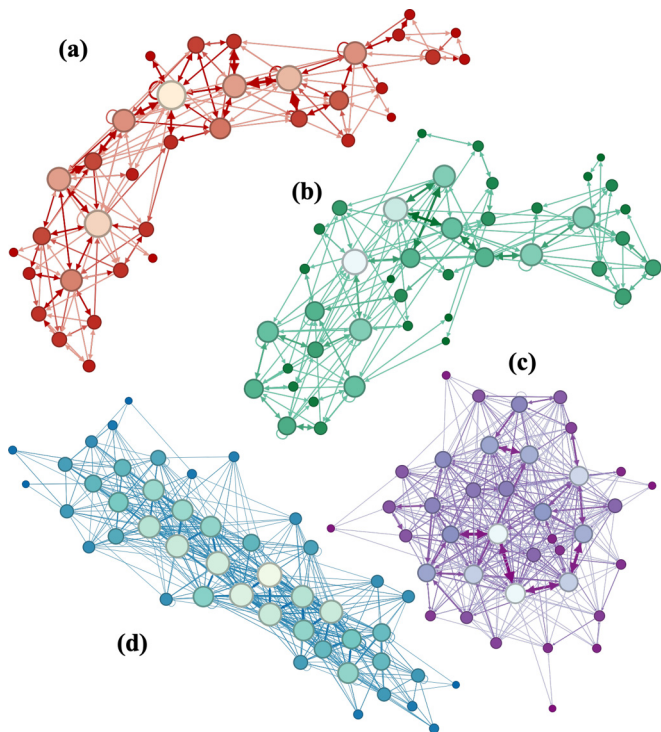


FIG. 2. Examples of note transition networks derived from select musical pieces: (a) Chorale BWV 437, (b) Fugue 11 BWV 856, from the *Well Tempered Clavier Book I*, (c) Prelude 9 BWV 878, from the *Well Tempered Clavier Book II*, and (d) Toccata BWV 916. Here we display the largest connected component of each network. The node size and color is based on the sum of its in and out degrees, such that nodes with a higher degree are larger in size and lighter in color, while nodes with a lower degree are smaller in size and darker in color. The edge thickness indicates the relative frequency of transitions.

Model information production using random walks

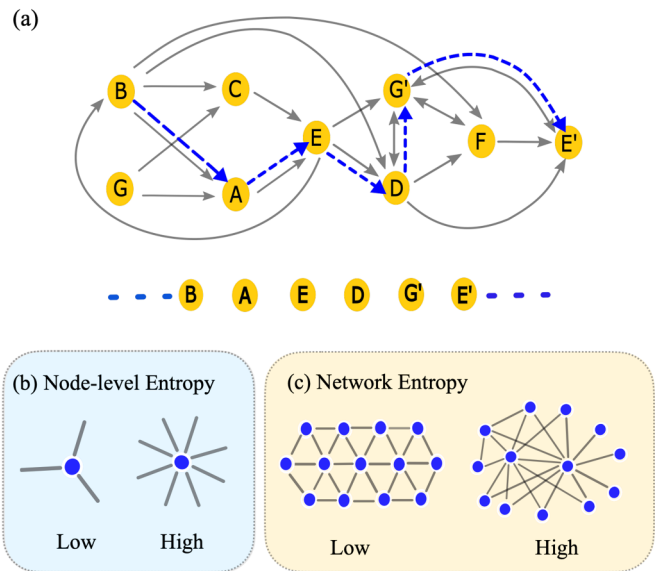


FIG. 3. The model of information production using random walks. (a) An example of a random walk on the network of note transitions is shown using the blue dotted line. At each node, the walker chooses an outgoing edge to traverse, each weighted with equal probability. This walk generates a sequence of notes as shown below. (b) The amount of information, or the *entropy*, generated when a walker traverses an edge from a node depends on the degree of the node. When traversing nodes with a high versus low degree, the walker has more choices for which edge to pick and, hence, such a transition generates more information. Thus, nodes with a higher degree (right) are said to have higher entropy than nodes with a low degree (left). (c) To calculate the entropy of the entire network, one needs to weigh the contribution of each node by the probability that a walker will occupy it. For networks with the same average degree, those with a wider range of degrees (right) have a higher entropy than those with a narrower range of degrees (left).

basic representation of the note transitions present in a musical piece, it is important to highlight the potential to extend this representation to capture other essential aspects of music (such as timbre, rhythm, duration of the notes, as well as techniques like counterpoint). We expand on how future efforts could incorporate more musical realism and complexity in Sec. VII.

III. QUANTIFYING THE INFORMATION IN NETWORKS

We seek to measure the amount of information produced by a sequence of notes. Although note sequences can have long-range temporal dependencies [52,53] and higher order structures [54,55], as a first analytical step, we focus on the Markov transition structure. That is, we study the information contained in individual note transitions. This information is quantified by the Shannon entropy of a random walk on the network [22,56] (Fig. 3; see also the Methods Sec. A 2 for further details). Given a network of transitions, the contribution of the *i*th node to the entropy can be written in terms of

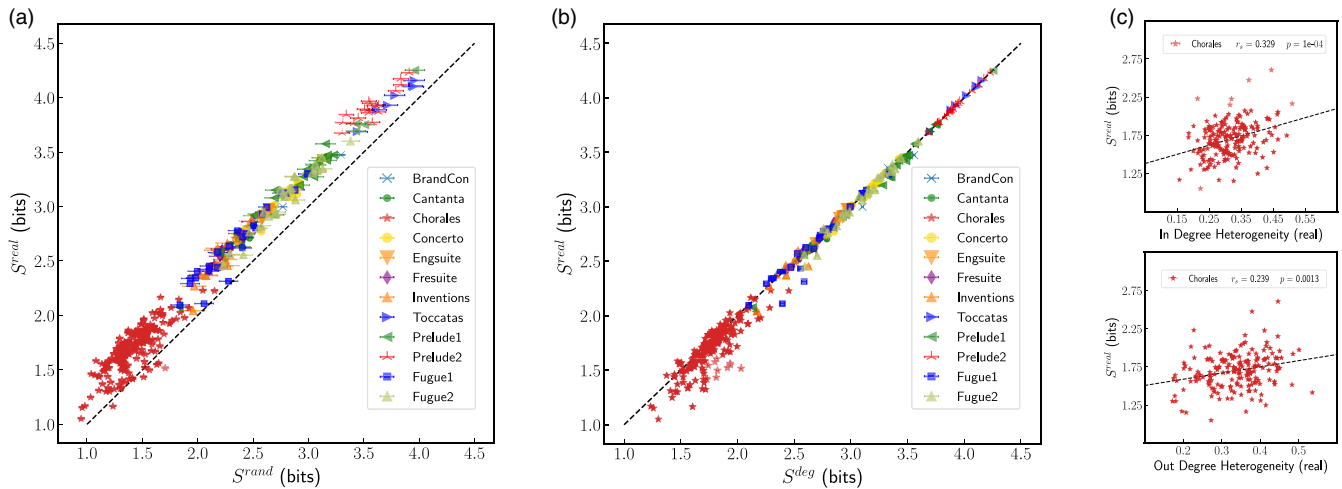


FIG. 4. Quantifying the information of Bach’s music using the entropy of random walks on networks of note transitions. (a) Entropy of Bach’s music networks (S^{real}) compared with random networks of the same size (S^{rand}). We report the entropy of the corresponding random networks after averaging over 100 independent realizations. The error bars for S^{rand} indicate the standard error of the sample. (b) The entropy of Bach’s music networks (S^{real}) compared with random networks that preserve the in and out degrees of each node (S^{deg}). We report the entropy of the corresponding degree-preserving random networks after averaging over 100 independent realizations. The error bars for S^{rand} indicate the standard error of the sample. (c) The entropy of the chorales as a function of the average in-degree heterogeneity $H^{\text{in}} = \text{Var}(k^{\text{in}})/\langle k^{\text{in}} \rangle$ (top) and out-degree heterogeneity $H^{\text{out}} = \text{Var}(k^{\text{out}})/\langle k^{\text{out}} \rangle$ (bottom) of the networks. In (a) and (b), each data point represents a single piece. Color and marker indicate the type of piece, as shown in the legend. The dashed line represents the line $y = x$. In (c), the dotted line indicates the best linear fit, the reported r_s value is the Spearman correlation coefficient, and the p value is obtained by performing a permutation test.

the entries of the transition probability matrix P as

$$S_i = - \sum_j P_{ij} \log P_{ij}. \quad (1)$$

In the case of directed unweighted networks, $P_{ij} = 1/k_i^{\text{out}}$, where k_i^{out} is the out degree of the node. Here, the base of the logarithm is 2. Hence, for unweighted networks, the node-level entropy is $S_i = \log(k_i^{\text{out}})$, which is solely determined by the out degree.

To calculate the entropy of the entire network, the contributions of the nodes are weighted by their stationary distribution—the probability that a walker ends up at node i after infinite time—which we denote by π_i [56]. The entropy of the network is then

$$S = \sum_i \pi_i S_i = - \sum_i \pi_i \sum_j P_{ij} \log P_{ij}. \quad (2)$$

For undirected and unweighted networks, the stationary distribution has a simple analytical form $\pi_i = k_i/2E$, where k_i is the degree of node i , and E is the total number of edges. The network entropy is then

$$S = \frac{1}{2E} \sum_i k_i \log k_i. \quad (3)$$

By contrast, for directed networks the stationary distribution depends on the detailed structure of the network and cannot be written in closed form. Hence, for our directed music networks, we calculate the stationary distribution numerically and use Eq. (2) to compute the entropy of each piece.

To understand the amount of information produced by the music networks, we compare them to randomized (or “null”) networks of the same size; that is, networks with the same number of nodes and edges (see Methods Sec. A 5

for details on generating null networks). This helps develop an intuition for the amount of information that networks of the same size typically contain. If the note transitions in the music networks do have distinct properties that allow them to communicate a large amount of information, then we would expect Bach’s networks to contain more information than the null transition structures. By averaging over 100 random networks for each piece, we find that the real networks generally have consistently higher entropy—thereby containing more information—than their random counterparts [Fig. 4(a)]. Moreover, by comparing across pieces, we observe that the different kinds of compositions cluster together based on their entropy. The chorales, typically meant to be sung by groups in ecclesiastical settings, are shorter and simpler diatonic pieces that display a markedly lower entropy than the rest of the compositions studied. By contrast, the toccatas, characterized by more complex chromatic sections that span a wider melodic range, have a much higher entropy. It is possible that the chorales’ functions of meditation, adoration, and supplication are best supported by predictability and hence low entropy, whereas the entertainment functions of the toccatas and preludes are best supported by unpredictability and hence high entropy. In the Supplemental Material [57], we perform comparisons for other information sources as well to provide more context to our result.

We know that the node-level entropy is defined only by the out degrees of the nodes. Accordingly, it is useful to assess differences between the true networks and others wherein the node-level entropies have been fixed by preserving the true degree distribution. To perform this assessment, we compare the entropy of the real networks with another set of null models: randomized networks which preserve both the in- and out-degree of each node (see Methods Sec. A 5 for details

on generating these networks). We observe that the entropies of the networks are more or less preserved [see Fig. 4(b)]. Although this preservation is expected for undirected networks (where the entropy is determined only by the degree distribution), it need not exist for directed networks (where the different stationary distributions contribute to the entropy). We therefore find that the entropy of music networks is primarily determined by their degree distributions rather than their stationary distributions.

To gain intuition for how the entropy of note transitions depends on network structure, consider the case of unweighted and undirected networks. The network entropy takes a particularly simple form, as shown in Eq. (3). Following a Taylor expansion around the average degree of the network (see Methods Sec. A 2), one obtains

$$S = \log\langle k \rangle + \frac{\text{Var}(k)}{2\langle k \rangle^2} + \dots, \quad (4)$$

where $\langle k \rangle$ is the average degree of the network and $\text{Var}(k)$ is the variance of the degrees. To first order, we see that the entropy increases logarithmically with the average degree of the network. To second order, the entropy increases with the variance or the *heterogeneity* of the degrees, such that more information will be produced by networks with heterogeneous (or broader) degree distributions. We define the degree heterogeneity as

$$H = \frac{\text{Var}(k)}{\langle k \rangle^2}. \quad (5)$$

Many networks that we encounter in our daily lives are characterized by heterogeneous degree distributions, typically with few high-degree “hub” nodes and many low-degree nodes [58–60]. By contrast, regular graphs—which have homogeneous degrees—produce random walks with the least entropy [see Fig. 3(c)].

Where does Bach’s music fall along this spectrum? We found in Fig. 4(a) that the music networks analyzed have consistently higher entropy than null networks with the same number of nodes and edges (in other words, randomized networks with the same average degree). In the Supplemental Material [57], we show that this higher information content of Bach’s music networks is due to higher heterogeneity in their in- and out-degree distribution; that is, the music networks are more heterogeneous in their degrees than expected from transition structures of their size, enabling them to pack more information into their structure. Since we have focused our analysis on the first-order sequential relationships among notes, which are likely common across different kinds of music, we expect this result to generalize for other kinds of music as well.

In Fig. 4(a), we also observe that various pieces belonging to certain compositional forms were clustered together in their entropy. Consistent with this observation, we find that the pieces which are clustered together in their entropy have very similar degrees (see Supplemental Material [57]). Examples include English suites, French suites, and chorales. In contrast, fugues did not cluster together in their entropy as much as other composition types and displayed diverse average degrees. For the compositions that are grouped together in their entropy, we find that the differences observed

among the pieces in the group can be explained by their degree heterogeneity (see Supplemental Material [57]). We can, for example, see this relation in the chorales where the pieces which have a higher in- and out-degree heterogeneity tend to have a higher entropy, despite having similar degrees [Fig. 4(c)]. To quantitatively verify this relationship, we compute the Spearman correlation coefficient between the network entropy and the in- and out-degree heterogeneity in Fig. 4(c). This coefficient assesses the monotonicity between two data sets, and ranges from -1 to 1 , with 0 implying no correlation. The positive correlation, as observed in the figure, suggests that the entropy tends to increase with in- and out-degree heterogeneity. For additional details on computing this correlation and the associated p value, refer to Methods Sec. A 7. Lastly, we note that this positive monotonic relationship between the entropy and degree heterogeneity holds even in our data set of *directed* networks, likely because the in- and out-degrees tend to be correlated.

IV. HOW HUMANS PERCEIVE NETWORKS OF INFORMATION

A key aspect of human communication involves receiving and assimilating information in the form of interconnected stimuli—ranging from sequences of words in language and literature to melodic notes of a musical piece, and even abstract concepts. Humans assimilate this information and build representations of the underlying structure of inter-item relationships, as depicted in Fig. 5(a). As noted earlier, humans build these internal network models using imperfect cognitive instruments that result in slightly distorted versions of true network structures. The information that is perceived by a human is the sum of the information present in the system and the inaccuracies that stem from the imperfect cognitive processes involved in perception [21]. In the previous section, we focused on quantifying the actual information present in the system (see Fig. 3). We will now account for the second piece: the inaccuracies that arise due to the imperfect cognitive process of perceiving information (see Fig. 5).

To understand how humans learn and represent transition structures, researchers have conducted a number of experiments and introduced a range of models describing how humans internally construct transition networks [35,36,49,61–63]. Several of these studies and models consistently highlighted a shared principle: Humans tend to integrate transition probabilities over time, relating items that are adjacent to each other as well as those separated by transitions of longer lengths—two, three, and beyond—with a declining likelihood as the transition distance increases [21,36,62,64]. Mathematically, we can capture this temporal integration and express the inferred transition structure \hat{P} in terms of the true transition structure P as

$$\hat{P} = \sum_{\Delta t=0}^{\infty} f(\Delta t) P^{\Delta t+1}, \quad (6)$$

where $f(\Delta t)$ is a decreasing function of Δt such that longer-distance associations contribute less to an individual’s network representation. This fuzzy temporal integration allows for lower computational costs and better generalizations about new information at the cost of accuracy. Here, we focus

Model perception using maximum entropy model

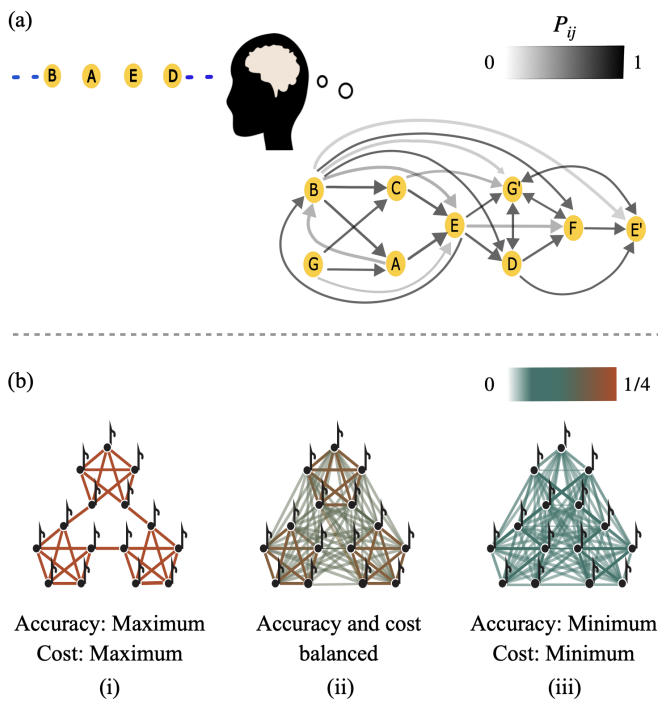


FIG. 5. How humans process networks of information. (a) A key aspect of human communication involves receiving and assimilating information in the form of interconnected stimuli. Humans assimilate patterns of information presented to them through imperfect perceptual systems, which results in slightly inaccurate internal models of the underlying transition structure. (b) When forming internal network models of the world, humans strike a balance between accuracy and complexity. The parameter η quantifies this trade-off between accuracy and cost. In panel (i), we see the example network built when solely maximizing the accuracy ($\eta \rightarrow 0$), which forms a perfect representation of reality. However, building this network requires perfect memory and is computationally expensive. In panel (iii), we see the network built when solely minimizing the computational cost ($\eta \rightarrow 1$), in which all nodes are connected to all other nodes, unlike the original network. Constructing this network does not require significant cost, but it provides no accuracy in representing the original information. Humans tend to display intermediate values of $\eta = 0.80$ [21], thereby constructing networks that preserve *some* but not all of the true transition structure, as shown in panel (ii). Figure adapted with permission from Ref. [51].

on one such model which captures this temporal integration and inaccuracies in perception [21,35]. The model postulates that when constructing internal network representations of information, humans aim to maximize the accuracy of their internal representation while simultaneously minimizing the computational cost required for its construction [21,35,51,65]. The learned transition probabilities (\hat{P}) can then be written in terms of the true transition probabilities (P) as follows:

$$\hat{P} = (1 - \eta)P(I - \eta P)^{-1}, \quad (7)$$

where $\eta \in [0, 1]$ captures the errors in representation. A detailed derivation of this expression is provided in Methods Sec. A.3.

To gain intuition for this model, it is useful to consider the two extreme limits of η . In the limit where $\eta \rightarrow 0$, the inferred network is exactly the same as the true network ($\hat{P} = P$). This scenario corresponds to that wherein the network is learned with no errors, forming an exact representation of the transitions [Fig. 5(b)(i)]. However, exactly learning the network is computationally expensive in that it requires perfect memory. By contrast, in the limit where $\eta \rightarrow 1$, the inferred network connects each node to every other node, and all of the structure is lost [Fig. 5(b)(iii)]. Such a representation is efficient to learn, but completely disregards accuracy. It turns out that most humans do something in between these two extremes by recalling the sequence of transitions sometimes accurately and sometimes inaccurately, thereby forming a fuzzy perception of the true network [Fig. 5(b)(ii)]. Formally, the competition between computational complexity and accuracy can be captured by a free-energy model of people's internal representation [35], and the inferred network under this model can be written in terms of the true network using Eq. (7). We emphasize the similarity of this form across multiple different theories of cognition [34,49,50]. By relating the inferred transition structure to the true network structure, this framework enables one to explore questions about the information that a human perceives from a given network. Given our interest in such questions in the context of music, we use this model to compute the inferred network for each musical piece. We emphasize that several empirical studies of musical expectancy have highlighted the role of statistical learning as a key mechanism, alongside other factors, in musical expectancy and knowledge acquisition [12,66–70].

For the rest of our discussion, we use the term “inferred network” on its own to refer to the network calculated using the model of perception discussed above. Prior work indicates that, on average, humans display an $\eta = 0.80$ in large-scale online laboratory experiments [21]. Given a network of note transitions with transition probabilities (P), we use this empirically measured value to calculate the inferred network (\hat{P}) using Eq. (7). In the context of music, it is important to recognize that the inferred structure would naturally exhibit variations, potentially influenced by factors such as an individual's level of training or cultural conditioning [41–44]. Nonetheless, this framework provides interesting insights regarding the types of structures that could be considered more effective in accurately communicating information, while taking into account the limitations of human perceptual systems. We provide a discussion of how future research could expand upon our research and improve the study of information perception in music in Sec. VII.

V. QUANTIFYING DISCREPANCIES IN THE PERCEPTION OF MUSIC NETWORKS

We are now prepared to investigate the extent to which the inferred music networks deviate from their true structure. Networks that display a low deviation between the inferred and true structure can be regarded as more effective in accurately communicating information. Hence, this framework provides insight into the communicative success of a network, from the point of view of how the network interacts with our imperfect perceptual systems. Mathematically, one can quantify the

deviations between the inferred network (\hat{P}) and the original network (P) using the Kullback-Leiber (KL) divergence:

$$D_{\text{KL}}(P||\hat{P}) = - \sum_i \pi_i \sum_j P_{ij} \log \frac{\hat{P}_{ij}}{P_{ij}}, \quad (8)$$

where π_i is the stationary distribution of the original network. The lower the KL-divergence, the closer the network is to the true network, and hence the network can be considered more effective in communicating information accurately. Do Bach’s musical compositions possess distinct features that result in smaller discrepancies in their perceived structure? How do pieces differ in these discrepancies? What are the structural differences between the musical pieces that lead to such differences?

To answer these questions, for each musical piece, we compute the KL divergence between the true transition probabilities P and the inferred transition probabilities \hat{P} . Then, to understand whether these music networks do indeed maintain low discrepancies in their inferred structure, we compare them against random networks with the same number of nodes and edges. The data confirms our intuition [Fig. 6(a)]: Bach’s music networks have a lower KL divergence than random networks of the same size. Even if we compare against null networks with the same in- and out-degree distributions, we still see that the music networks have a lower KL divergence [Fig. 6(b)]. This finding suggests that the lower KL divergence of these networks cannot be explained by their degree distributions alone. Additionally, we observe interesting variations in the KL divergence among the different compositional forms (Fig. 6). The chorales, at one extreme, seem to have the highest KL divergence, while the preludes and toccatas have the lowest KL divergence. In what follows, we attempt to identify and interpret the network properties that underlie the observed variations in the discrepancies of the inferred information across compositional forms and pieces.

A. Transitive clustering coefficient

As seen in the previous section, the discrepancies in the inferred transition structure for the music networks could not be explained by the distribution of degrees alone. For undirected networks, prior research has demonstrated that the KL divergence between the inferred and true transition structures decreases with an increase in the density of triangles within the network [21]. This relationship can be demonstrated by substituting the expression for the inferred version of a network [Eq. (7)] into the equation for the KL divergence [Eq. (8)]. We now extend this analysis to our *directed* networks, with the aim of generalizing this finding. By performing this substitution, we derive the subsequent expression for the KL divergence in terms of the original network’s adjacency matrix (A):

$$D_{\text{KL}}(P||\hat{P}) = - \log(1 - \eta) - \frac{\eta}{\ln 2} \sum_i \pi_i \times \left\{ \sum_j A_{ij} \sum_l \frac{1}{k_i^{\text{out}}} A_{il} \frac{1}{k_l^{\text{out}}} A_{lj} \right\} + O(\eta^2). \quad (9)$$

Here we see that the KL divergence depends on a product of the form $A_{ij}A_{il}A_{lj}$, which quantifies the *transitive* relationships present in the network. More explicitly, it depends on the number of directed triangles of the form $i \rightarrow j \rightarrow k$ and $i \rightarrow k$.

To quantify the extent to which a network has clusters of this form, we introduce a measure termed the *transitive clustering coefficient* of the network, defined along similar lines to the clustering coefficient of a network [71,72]. For each node, this quantity is measured by dividing the number of transitive triangles that node i is a part of (Δ_i^T) by the number of possible directed triangles:

$$C_i^T = \frac{\Delta_i^T}{k_i^{\text{tot}}(k_i^{\text{tot}} - 1)}. \quad (10)$$

Here k_i^{tot} is the total degree (in + out) of the node. We average this quantity over all nodes in the network to report a single value for each piece. As indicated by Eq. (9), we expect the KL divergence of the networks to primarily be driven by the transitive clustering coefficient. This relationship is indeed evident in Fig. 6(c), where we observe that musical networks with a higher transitive clustering coefficient tend to exhibit lower KL-divergence values. In this context, we also observe that the preludes and toccatas (which demonstrated relatively lower KL-divergence values) are characterized by a larger density of transitive triangles compared to other pieces like the chorales.

A natural question that arises at this point is, What is the significance of these transitive relationships within the networks, and why do they contribute to reduced disparities between the inferred and true structure? From a cognitive science perspective, this relationship between the KL divergence and clustering arises from the tendency of humans to count transitions of length two, as discussed previously. In a scenario where a given node i is connected to node j and node j links to node k , a human learner may erroneously draw an edge between node i and node k in their mind. However, if the network originally had a direct link from node i to node k , such an error would reinforce an existing edge, thereby aligning the inferred network more closely with the true network. Hence, we expect networks with high clustering to be more robust to errors made during inference. From a music perspective, interpreting these triangles is not straightforward since the networks are unweighted. Nevertheless, the presence of a large density of such triangles suggests that if there is a transition between notes i and j and notes i and k , there is likely also a transition between notes j and k . This could potentially reflect the tendency of music to form tonally stable sequences of note transitions. Substantiating these claims would require further efforts, which we elaborate on in Sec. VII.

Analyzing the transitive clustering further, we find that the musical networks have a higher transitive clustering coefficient than degree-preserving random networks [Fig. 6(d)], suggesting that this feature is not due to mere coincidence. From Fig 6(d), we make an interesting observation: the preludes appear to have a *lower* transitive clustering coefficient than the corresponding null networks that preserve their size and degree distribution, while the chorale pieces generally

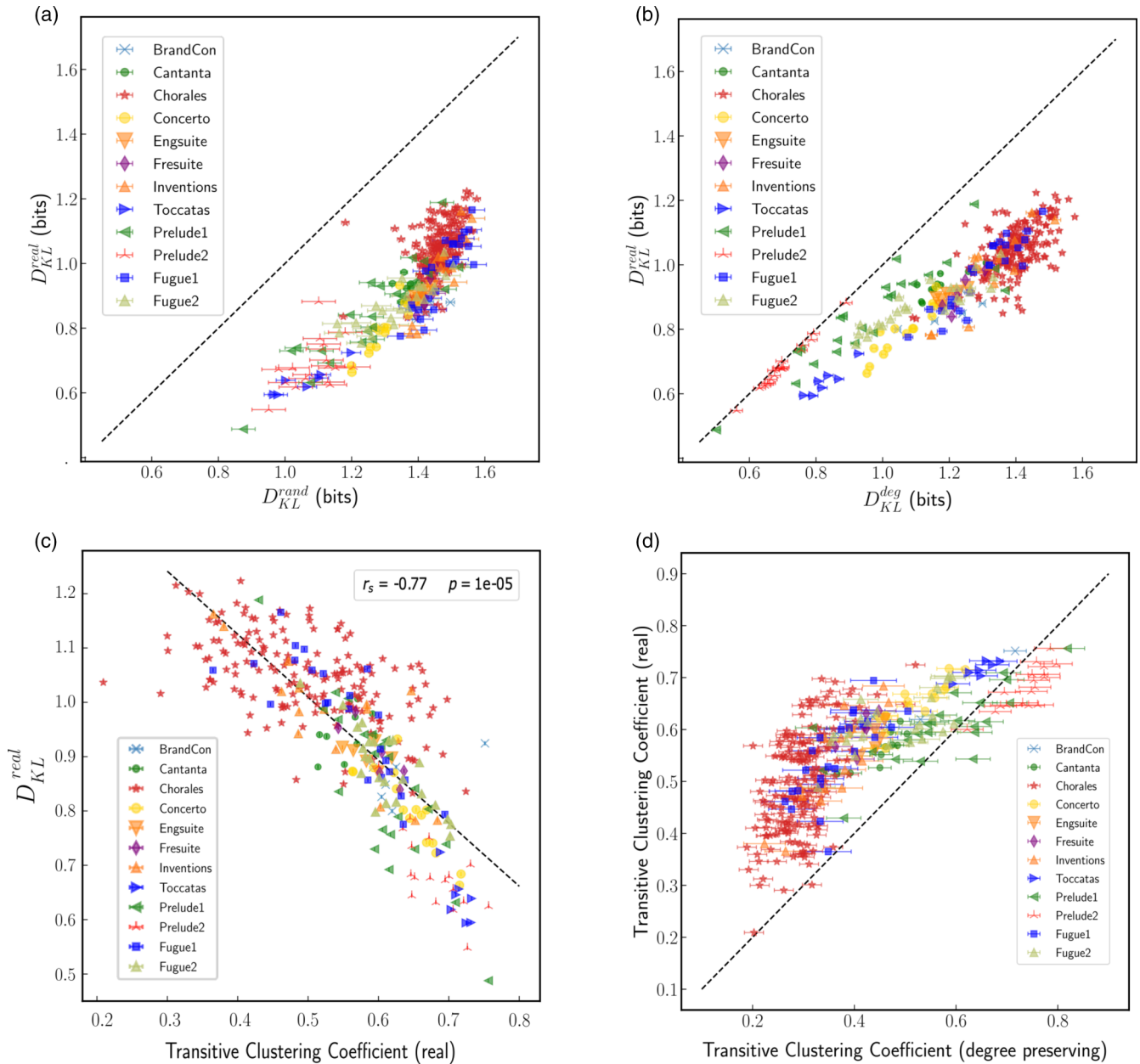


FIG. 6. Quantifying the difference between the actual information and the perceived information in Bach’s music networks by calculating the KL-divergence between the actual and perceived network. (a) KL divergence of the real music networks (D_{KL}^{real}) compared with random networks of the same size (D_{KL}^{rand}). We report the KL divergence of the corresponding random networks after averaging over 100 independent realizations. The error bars for D_{KL}^{rand} indicate the standard error of the sample. (b) KL divergence of the real music networks (D_{KL}^{real}) compared with random networks that preserve the in and out degrees of each node (D_{KL}^{deg}). We report the KL divergence of the corresponding degree-preserving random networks after averaging over 100 independent realizations. The error bars for D_{KL}^{deg} indicate the standard error of the sample. (c) KL divergence of the real music networks as a function of the transitive clustering coefficient of the network $C = \langle \Delta_i^T / k_i^{tot} (k_i^{tot} - 1) \rangle$. (d) The transitive clustering coefficient of the real music networks compared with random networks that preserve the in and out degrees of each node. The dotted line indicates the line $y = x$. For the degree-preserving random networks, we report the transitive clustering coefficient after averaging over 100 independent realizations, with error bars denoting the standard error of the sample. In all panels, each data point represents a single piece. Color and marker indicate the type of piece, as shown in the legend. The dotted lines in panels (a), (b), and (d) represent the line $y = x$.

have a higher transitive clustering coefficient than expected from null networks. We probe this further in the Supplemental Material [57] and identify mesoscale structures that could lead to the observed differences between the compositional forms.

VI. ACCOUNTING FOR NOTE TRANSITION FREQUENCIES

So far, we have focused our attention on the information content and perception of unweighted (or binary) note

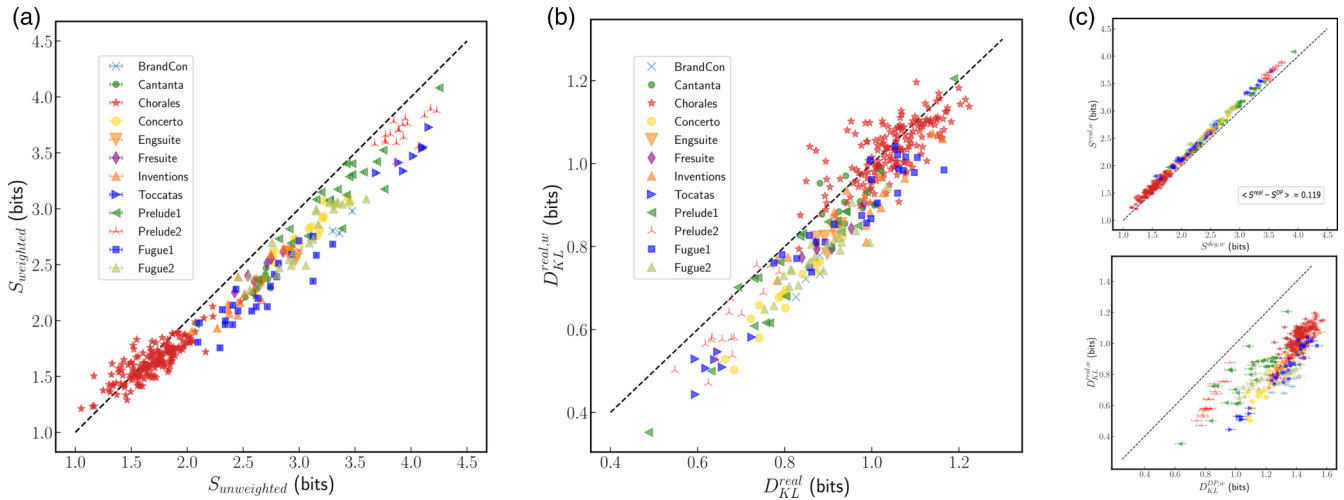


FIG. 7. Accounting for the frequencies of the note transitions in our analysis. (a) Entropy of the weighted versions of Bach’s music networks (S_{weighted}) compared with the corresponding unweighted versions ($S_{\text{unweighted}}$). (b) The KL divergence of the weighted versions of Bach’s music networks ($D_{\text{KL}}^{\text{real},w}$) compared with the corresponding unweighted versions ($D_{\text{KL}}^{\text{real}}$). (c) Top: Entropy of the weighted note transition networks ($S^{\text{real},w}$) compared with degree-preserving edge-rewired null networks ($S^{\text{deg},w}$). Bottom: The KL divergence of the weighted note transition networks ($D_{\text{KL}}^{\text{real},w}$) compared with degree-preserving edge-rewired null networks ($D_{\text{KL}}^{\text{deg},w}$). In all panels, each data point represents a single piece. Color and marker indicate the type of piece, as shown in the legend. The dashed line represents the line $y = x$. In the top figure of panel (c), we report the average deviation of the data points from the line $y = x$.

transition networks created from Bach’s music. These networks only captured whether or not a transition exists between two notes and were not sensitive to how frequently each transition occurs. The binary networks enabled us to probe how the structure of the transitions supports effective communication. However, in many real networks, not all transitions occur with the same frequency. To reflect the different frequencies with which transitions may occur, we construct networks in which transitions are weighted according to this. For example, if note i follows note j 90% of the time and note k follows note j 10% of the time, the edge from node j to node i will be more heavily weighted than the edge from node j to node k (see Methods Sec. A 1 for further details on network construction). Adding this piece of information to the networks leads us to new questions about the role that transition weights play in communicating information to listeners. For example, how is the information generated by a random walk on the network altered by differences in the frequencies of transitions? Do these differences in frequencies reduce the discrepancies in the inferred network?

A. Weights reduce the surprisal of transitions

For unweighted networks, the node-level entropy of a random walk is determined solely by the out degree (k_i^{out}), since each outgoing edge is traversed with probability $P_{ij} = 1/k_i^{\text{out}}$. If the edges are weighted by their transition frequencies, the P_{ij} ’s will no longer be uniformly distributed, and each outgoing edge will not have an equal probability of being traversed. Hence, incorporating the edge weights reduces the node-level entropy. This observation is intuitive since nonuniformities in any distribution lead to decreases in entropy. However, extending this intuition to the entropy produced by the entire network is not as straightforward, since one must weigh the contribution of each node by

the stationary distribution of the random walkers, which cannot be expressed in closed form for directed networks. Generally, we find that the entropy of weighted networks is still *lower* than the corresponding unweighted networks [Fig. 7(a)]. This finding suggests that the different weights do indeed reduce the overall surprisal generated by the networks.

B. Weights reduce discrepancies between the inferred network and the original network

Incorporating the transition frequencies also helps us to understand the role that the weights play in the human inference of note transitions. We observe that the weighted networks of note transitions have *lower* KL divergence than the binary networks [Fig. 7(b)]. This observation suggests that the weights aid in forming more accurate internal representations of the transition structures, thereby reducing the discrepancies between the inferred and true structure.

In light of these data, we next verify the role that the network structure plays in the communicative success of weighted networks by comparing the entropy and KL divergence of the weighted music networks with edge-rewired null networks. In the analysis on unweighted networks, we observed that the entropy was primarily driven by the degree distribution of the network and not sensitive to the precise connectivity pattern. To make this observation, we compared the entropy of the real music networks to randomized networks that preserved the exact degree distribution of each node and, hence, held the node-level entropies fixed. Along similar lines, here we make use of null models that keep the node-level entropies fixed by preserving the in and out degrees of each node *and* the outweights at each node (see the Methods section for details on the null models). By comparing the entropy of the weighted music networks to

the degree-preserving weighted null models, we see that the entropies of real networks are still more or less unchanged, although the real networks have marginally higher entropies than the null networks [Fig. 7(c), top]. These results support our conclusion that the entropy in the real networks is still primarily driven by their degree distribution. When we compare the KL divergence of the real weighted networks with the degree-preserving weighted null models, we find that the real networks have a lower KL divergence than the corresponding null networks [Fig. 7(c), bottom]. Together, these results suggest that incorporating the weights into our network analysis does not alter our results on the effects of network structure qualitatively.

Accounting for the note transition frequencies in our network model leads to several interesting lines of inquiry. For instance, is it the specific distribution of weights that improves the accuracy of the inferred music networks? Future work could evaluate this possibility by comparing the KL divergence of the weighted networks with a class of null models that preserve the skeleton of the network but permute the edge weights. It would also be interesting to test whether higher edge weights are concentrated in triangular clusters of the network, offering a potential explanation for the lower KL divergence of the weighted networks compared to the binary networks. Since the weights correspond to the number of times a transition occurs, we anticipate that analyzing the distribution of these edge weights would give us information about the tonal characteristics of a musical piece. Atonal music, in particular, can be expected to exhibit a lower heterogeneity in the edge weights. Such a quantitative approach to assessing the tonal characteristics of a piece ties into existing and ongoing research on tonality in musical compositions as well as the perception of tonality [73–76]. We elaborate on how our framework serves as a platform for further research in this area in Sec. VII.

VII. CONCLUSIONS AND FUTURE DIRECTIONS

Across language, literature, music and even abstract concepts, humans demonstrate the remarkable ability to identify patterns and relationships from sequences of items—an essential aspect of information sharing and communication [35,51,77–79]. Here, we draw upon ideas from network science, information theory, and statistical physics to build a framework that serves as a stepping stone for studying the information conveyed by a musical piece. We use this framework to analyze networks of note transitions in a wide range of music composed by J. S. Bach. For each musical piece, we construct a network of note transitions by drawing directed edges between notes that are played consecutively. We then quantify the amount of information generated by the network structure and find that different compositional forms can be grouped together based on their information entropy. We relate the information content of each piece to its network structure, enabling us to gain insight into the structural properties of various pieces. Next, inspired by recent progress in the field of statistical learning which demonstrates how humans infer transition structures across visual and auditory domains [35,66,67,78], we use a computational model [21,35] for how humans learn networks of information to compute

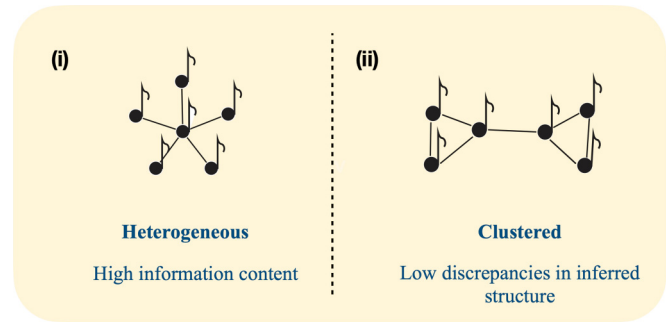


FIG. 8. Network structures that support effective communication of information. Networks with a larger variance or heterogeneity in their node degrees, as shown in panel (i), pack more information into their structure and have a higher entropy. Clustering in the network, as shown in panel (ii), makes the structure more resilient to errors made by humans when building an internal representation of the information, allowing the network to be inferred more accurately. Together, these structures convey a large amount of information that can be learned by humans more accurately, and are hence more efficient for communication.

the average “inferred” network structure for each piece. We then quantify the discrepancies between the inferred and true transition structures under this model. Here too, we observe interesting differences among the pieces, which we attribute to differences in the clustering of the networks. Finally, we study how the frequencies of transitions influence the information content and perception of the musical pieces by weighing the transitions by the number of times they occur. We find that the weights reduce the overall entropy or surprisal of the transitions, and also reduce the deviations between the inferred and actual network, suggesting that the weights aid in accurate inference of these transition structures.

Furthermore, we find that the music networks contain more information and maintain lower discrepancies in the inferred structure than expected from typical transition structures of the same size. This provides us insight into features that make networks of information effective at communicating information. In general, networks which are denser (have a higher average degree) produce more information (have a higher information entropy). For networks of comparable average degree, more heterogeneous (higher variance in degree distribution) structures produce more information than those that are more regular or homogeneous in their degree [Fig. 8(i)]. Moreover, networks which have a high degree of clustering maintain a lower divergence from human expectations [Fig. 8(ii)]. Together, these findings suggest that for networks of a given size, rapid and accurate communication of information is supported by structures that are simultaneously heterogeneous and clustered (Fig. 8). Notably, such structures are widely prevalent across complex systems [58–60,71,80].

We hope that our framework inspires further exchange between physics, cognitive science, and musicology. On a broader scale, our study also adds to investigations on how information in complex systems is structured. To conclude, we highlight a number of exciting directions for future inquiry and outline ways in which our framework can be expanded upon and improved.

A. Future directions

A natural follow-up to this analysis would be to examine works of other composers—particularly works outside the Western tradition. This also prompts questions aimed at assessing how various styles or genres of music differ [81–83]. In particular, what are the key features by which a listener distinguishes between music from two eras, say the Classical and the Romantic eras? How do the differences in structure then impact how the piece is perceived by a listener? Consequentially, a quantitative assessment of musical compositions like ours raises the intriguing possibility of identifying works of a composer or genres that may not be *a priori* obvious to musicologists.

Systematically analyzing the information that we extract from complex systems also provides us with new tools to understand human creativity and experiences. A question that often arises in the context of how humans experience music is, What makes a musical composition appealing to the human ear? While individual preferences in music can vary widely and are highly subjectively, there is still a general agreement on certain composers being considered influential or great. This fact raises the possibility that there may be some inherent qualities that are common to musical pieces which are widely considered appealing. Identifying such features might give us insight into the creative process of composing music and also complement existing work using AI to generate music [84,85]. Several attempts have been made to identify such patterns. For example, Ref. [19] analyzed note transition networks in certain compositions by Bach, Chopin, and Mozart as well as Chinese pop music, and suggested that “good” music is characterized by the small-world property [71] and heavy-tailed degree distributions. On the other hand, Ref. [30] studied selected compositions from Bach’s *Well-Tempered Clavier* and found non-heavy-tailed degree distributions, suggesting that such distributions are not necessary for music to be appealing. It would be interesting to devise future experiments to determine whether our findings relate to the aesthetic or emotional appeal of a piece. In our study, we found that Bach’s music networks had a higher number of transitive triangular clusters, enabling them to be learned more efficiently than arbitrary transition structures. Are pieces with a larger number of these triangles also more appealing to a listener? Future work could assess this possibility by conducting experiments that ask people to rate Bach’s compositions and analyzing whether these ratings correlate with the presence of triangular clusters.

Additionally, one could explore how the tonal patterns and characteristics of a musical composition influence its general aesthetic appeal. As suggested in Sec. VI, examining the distribution of edge weights might give us information about the tonal characteristics of a musical piece. Conducting experiments to explore the relationship between the diversity of edge weights in a music network and the overall aesthetic appeal of note sequences generated by the network would be an interesting avenue for future work. Such a quantitative approach would complement the current and ongoing studies in this area [73–75]. More generally, our work focuses not solely on the information inherent in the transition structure of music but also on how the information in this transition

structure is perceived by a human listener. This framework might be useful in studying cognitive aspects of music and in bridging patterns observed in data with cognitive theories of music.

In future work, it also would be interesting to extend our analysis to examine how music networks evolve with time. There are three potentially interesting lines of inquiry here: First, how do the entropy and KL divergence of a musical piece change as the piece progresses? Does this temporal change differ among the various compositional forms? Second, how has the music of a specific composer (whether Bach or otherwise) changed over the course of their lifetime? Has it become more intricate and complex, holding more information? Perhaps as the composer gains experience, their compositions convey information more efficiently and accurately, as reflected in a reduced KL divergence? If the exact dates of when each piece was composed were known, then the framework used in our paper might provide answers to these questions. Third, how has music of a given genre, say classical music, changed over the years across composers? Reference [32], for example, studied the fluctuation in pitch between adjacent notes in compositions by Bach, Mozart, Beethoven, Mendelsohn, and Chopin, and found that the largest pitch fluctuations of a composer gradually increased over time from Bach to Chopin. As mentioned earlier, it would be interesting to expand our analysis to different composers, and see how the information and expectations vary across composers and time. Tracing how other quantitative network characteristics (for example, the presence of tonal patterns, as indicated in Sec. VI) have evolved over time across different music traditions would also be another avenue for further study.

Finally, the quantitative measures described in our work could serve as feedback quantities to human composers to aid their composition process. For example, in music composition software, one could add a feature that displays the music network composed thus far and its entropy measures, etc. Composers can then use these measures as feedback in their music writing process. Using our framework, the software could even suggest edits that could alter the surprisal—potentially suggesting ways to contradict musical expectation (resulting in a higher entropy) or harmoniously resolve notes to meet those expectations. Moreover, capturing data from users as they compose and employing our quantitative framework would add to research mapping out the creative process of musical composition [86,87]. This approach goes beyond analyzing only the final product, therefore providing deeper insight into the workings of human creativity.

Before we conclude, we highlight limitations within our analysis that highlight directions for further effort. First, our paper relies on a simplistic representation of music that could be expanded to incorporate more musical realism and complexity. For instance, one could account for differences in timbre, the intervals between notes, or even fused notes or chords, which are known to play a key role in music perception [88,89]. Second, while we have focused on the information present in first-order sequential relationships among the notes, future work could capture higher-order correlations, hierarchies, and more intricate structures inherent in music [52–55,90,91]. Recent advances in studying higher-

order dependencies and structures present in networks offer a promising approach to capturing this complexity [92–94]. An essential feature to capture in the analysis of polyphonic music is counterpoint, or in other words, the presence of relationships between distinct lines of music. Our current analysis computes the note transitions separately for different lines and combines them to form a single network, and thereby does not capture counterpoint. It is important to capture this (among other techniques used in music) in future work. One potential method to incorporate this complexity is by extending our current method using multilayer networks [95–98]. In this framework, the note transition networks derived from distinct musical lines would constitute distinct “layers” within the multilayer network. Then, the relationships across lines (or layers) could be captured through interlayer edges, potentially quantified by measuring correlations across musical lines. The process of information generation could then be modeled using a coupled process, each running on a different layer but coupled via the interlayer edges [99]. A similar multilayer network framework could be used to simultaneously capture other aspects of music—such as rhythm or timbre—in addition to note transitions.

Incorporating such subtleties would not only improve our understanding of how the networks are structured, but also how they are perceived. Expanding on this understanding, it would be beneficial to conduct targeted experiments that specifically address and build models of the perception of distinct (but interdependent) musical attributes. Further, exploring the variability of music perception among individuals, considering factors such as musical training or cultural influences would also be interesting. We highlight existing research that investigates deviations from expectation in music based on such nuances. Notably, studies referenced in Refs. [41–44] suggest that cultural knowledge shapes musical expectancies and perceptions of musical complexity, particularly when individuals encounter novel musical forms. Similarly, research in Ref. [100] illustrates that even rhythm perception can significantly depend on culture. Investigations in Refs. [45–47] propose that language background, particularly background in a tonal language, impacts the perception of musical sequences. Moreover, the level and type of musical training are also known to influence music perception [37–40]. While our present framework and computational model does not account for these nuances, it provides a benchmark for comparative analysis across cultures and musical training levels that could separate out elements that are universally easier to anticipate, leaving room for subsequent research to delve deeper into how the aforementioned specific factors modify these general predictive mechanisms. Naturally, expanding computational models to encompass these nuanced factors is a direction for further inquiry.

The aforementioned and ensuing directions would expand our capacity to address more specific questions regarding the composer idiosyncrasies, era characteristics, and genre discussed earlier. As such, our work offers a flexible framework that can be utilized by a wide range of scholars both in and outside of physics. Beyond music, our study can also be extended to a range of complex systems present around us—such as language and social networks. For example, one

could analyze works of literature and ask, Does the entropy of noun transitions in various works of Shakespeare differ based on their genre? More specifically, does the information content and learnability of noun transitions or relationships between characters differ between tragedies and comedies? By providing an example of a systematic and comprehensive analysis of the actual and perceived information in music, our study complements and adds to the rich study of language, music, and art as complex systems [30,101,102]. Finally, a quantitative treatment of the patterns and motifs present in music complements research exploring analogies between music and other fields of science—including understanding protein structures and designing materials [103–105].

The data and code used to perform the analyses in this paper are openly available at Ref. [106].

ACKNOWLEDGMENTS

We thank C. Macklin for an early conversation on this topic and audience members who asked probing questions about our earlier work in communication networks. These interactions motivated our continued investigation in this space. We would also like to thank S. Varkey, S. Patankar, and A. Winn for useful conversations and feedback on earlier versions of this paper. This particular research was primarily supported by the Army Research Office Award No. DCIST-W911NF-17-2-0181 and the National Institutes of Mental Health Award No. 1-R21-MH-124121-01. D.S.B. would also like to acknowledge additional support from the John D. and Catherine T. MacArthur Foundation, the Alfred P. Sloan Foundation, the Institute for Scientific Interchange Foundation, and the Army Research Office (Grafton-W911NF-16-1-0474). The content is solely the responsibility of the authors and does not necessarily represent the official views of any of the funding agencies.

APPENDIX A: FURTHER DETAILS ON DATA AND METHODS

1. Data collection and network construction

The music files were collected in the MIDI format from various sources. The sources for the compositions analyzed are as follows: preludes [107,108], fugues [107,108], inventions [107,108], cantatas [109], English suites [110], French suites [110], four-part chorales [108], Brandenburg concertos [108], toccatas [110], and concertos [110]. The preludes and fugues are split based on whether they belong to the first or second part of *The Well-Tempered Clavier*, and are labeled 1 or 2. Certain compositions consist of different movements and our data set has separate MIDI files for each movement. We analyze each movement separately and average our measurements over them to yield a single measured quantity for each piece, as indexed by a unique Bach–Werke–Verzeichnis (BWV) number. In the case of the chorales, we analyzed the 186 four-part chorales in Bach-Gesellschaft Ausgabe (BGA) Vol. 39 with BWV No. 253-438.

The MIDI files were read in MATLAB using the `readmidi` function in MATLAB [111] to obtain information about the notes being played. Different instruments in a piece are stored

in separate channels within each data file. The transitions between notes are calculated separately for each instrument or track. We assign each note present in a piece a node in the network, and notes from different octaves are assigned distinct nodes. We then draw an edge from note i to note j if there is a transition between them. If there are multiple notes being played at a single time t (as is the case with chords), edges are drawn from the previously played note to *all* notes at time t , and from all the notes being played at time t to the subsequent note(s). This procedure gives us a directed binary network of note transitions. Finally, we restrict our analysis to the largest connected component of each network. The code and data used to construct the networks is available at Ref. [112]. We also construct weighted versions of these networks, where each edge is weighted by the number of times the corresponding transition occurs.

2. Entropy of random walks on networks

We use random walks to model how a sequence of information is generated from an underlying network of information. Under this model, a walker traverses the network by picking an outgoing edge to traverse at each node. Given a network with adjacency matrix A and matrix element A_{ij} , the probability that a walker transitions from node i to node j in a standard Markov random walk is $P_{ij} = A_{ij}/k_i^{\text{out}}$, where $k_i^{\text{out}} = \sum_j G_{ij}$ is the out degree of a node. We are interested in quantifying how much information is contained in the resulting sequence, which is captured by the entropy of the random walk:

$$S = - \sum_i \pi_i \sum_j P_{ij} \log P_{ij},$$

where π is the stationary distribution of the walkers, which satisfies the condition $P\pi = \pi$. For the simplest possible case of an undirected and unweighted network, $P_{ij} = 1/k_i$ and $\pi_i = k_i/2E$, where k_i is the degree of the i th node and $E = \sum_{i,j} A_{ij}/2$ is the total number of edges. The entropy in this case simplifies to

$$S = \frac{1}{2E} \sum_i k_i \log k_i = \frac{\langle k \log k \rangle}{\langle k \rangle}. \quad (\text{A1})$$

We can apply a Taylor expansion to this expression around the average degree of the network, and thereby obtain

$$S = \log \langle k \rangle + \frac{\text{Var}(k)}{2 \langle k \rangle^2} + \dots \quad (\text{A2})$$

Hence we find that the entropy of random walks increases logarithmically with the average degree of the network. Additionally, it grows as the variance of the degrees increases. This formalization enables us to relate the information content of various music networks to their network structure. The code used to measure the entropy of random walks on the networks analyzed is available at Ref. [112].

3. Model for how humans learn networks

As discussed in the main text, humans do not infer the transition probabilities of sequences of information with perfect accuracy due to imperfections in their cognitive processes.

Studies have consistently found that in forming internal representations of transition structures, humans integrate transition probabilities over time [21,36,62,64]. This process results in humans connecting items in the sequence that are not directly adjacent to each other. Mathematically, we can express the inferred transition structure \hat{P} in terms of the true transition structure P under this model of fuzzy temporal integration as

$$\hat{P} = \sum_{\Delta t=0}^{\infty} f(\Delta t) P^{\Delta t+1}, \quad (\text{A3})$$

where $f(\Delta t)$ is the weight given to the higher powers of P and is a decreasing function of Δt such that longer-distance associations contribute less to a person's network representation. The functional form of $f(\Delta t)$ is obtained using the free-energy model described in Ref. [35]. This model suggests that when forming internal representations of information, each human arbitrates a trade-off between accuracy and cost. The optimal distribution for $f(\Delta t)$ under this model is then a Boltzmann distribution with a parameter β that quantifies the trade-off between cost and accuracy in forming an internal representation of the information,

$$f(\Delta t) = e^{-\beta \Delta t} / Z, \quad (\text{A4})$$

where $Z = \sum e^{-\beta \Delta t} = (1 - e^{-\beta})^{-1}$ is a normalization constant. Substituting this expression to simplify Eq. (A3), we obtain an equation that relates the inferred transition probabilities \hat{P} to the true transition probabilities P ,

$$\begin{aligned} \hat{P} &= (1 - e^{-\beta})^{-1} \sum_{\Delta t=0}^{\infty} e^{-\beta \Delta t} P^{\Delta t+1} \\ &= (1 - \eta) P (I - \eta P)^{-1}, \end{aligned} \quad (\text{A5})$$

where $\eta = e^{-\beta}$. Prior work has estimated the value of η to be 0.8 from large-scale online experiments in humans [21]. Using this measured value of η , we use Eq. (A5) to calculate the inferred network for any given music network (code available at Ref. [112]).

4. KL divergence

To quantify how much the distorted learned transition structure \hat{P} differs from the original transition structure P , we calculate the KL divergence between the two transition structures. The KL divergence is a measure of how different a probability distribution is from a reference distribution and is given by

$$D_{\text{KL}}(P||\hat{P}) = - \sum_i \pi_i \sum_j P_{ij} \log \frac{\hat{P}_{ij}}{P_{ij}}, \quad (\text{A6})$$

where $\bar{\pi}$ is the stationary probability distribution of the transition matrix P , obtained by solving $P\pi = \pi$. The KL divergence between two quantities is always non-negative and attains the value zero if and only if $P = \hat{P}$. The larger the KL divergence, the more the inferred network \hat{P} differs from the original network. Hence, this quantity acts as a measure of the extent to which a network gets scrambled by the inaccuracies of human learning—or in other words, how accurately the network structure is inferred.

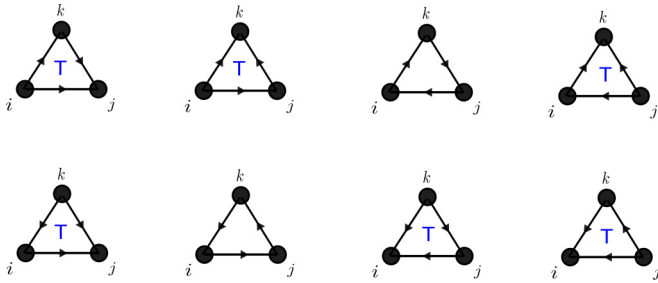


FIG. 9. The eight different possible triangles with node i as a vertex in a directed graph. The triangles which represent transitive relationships are marked using the letter T.

5. Null models

We aim to identify distinct features in the music networks that enable them to convey information effectively. To assess whether our observations are merely due to random chance or are instead a unique feature of our data set, we compare our measurements on the real music networks with the following null network models [113,114]:

(1) *Null networks with the same number of nodes and edges.* These are obtained by generating random networks with the same number of nodes and edges, and enable us to assess whether the quantity we have measured is to be expected merely based on network size.

(2) *Degree-preserving null networks.* These are randomized networks of the same size, with the additional constraint that the in and out degrees of *each* node in the network are preserved. Such networks are constructed by swapping edges between pairs of nodes in the network iteratively, such that the in and out degrees of each node are preserved but the connectivity (or topology) of the network is randomized. This class of null models enable us to evaluate the role that connectivity or topology plays in the quantity we are measuring.

We can generalize the degree-preserving null networks to weighted networks. We are interested in degree-preserving randomized networks since these keep the node-level entropies fixed and allow us to study the impact of topology on the quantities we are measuring. In the case of weighted networks, the node-level entropies are determined by the out weights and out degrees of the nodes. Hence, our procedure of swapping edges between pairs of nodes in the network still works since it preserved the out weights of each node in addition to the in and out degrees. With these null models, we can benchmark the presence of the quantities we are interested in and identify the role that the connectivity pattern or size plays. The code used to generate the null networks is available at Ref. [112].

6. Transitive clustering coefficient

Along the lines of the clustering coefficient of a node [71,72], we define the transitive clustering coefficient as a measure of the degree to which nodes in a directed network tend to form transitive relationships. The transitive clustering coefficient of a node i (for an unweighted graph with no self loops) is given by

$$C_i^T = \frac{\Delta_i^T}{k_i^{\text{tot}}(k_i^{\text{tot}} - 1)}, \quad (\text{A7})$$

where Δ_i^T denotes the number of transitive triangles that node i is a part of and k_i^{tot} is the total degree (in + out) of the node. The denominator simply counts the number of triangles that could exist within the neighborhood of node i .

The possible directed triangles involving node i can be divided into two categories—those representing cyclic relationships and those representing transitive relationships (Fig. 9). The number of transitive triangles involving node i that actually exist can be expressed in terms of the adjacency matrix of graph A :

$$C_i^T = \frac{(A + A^T)_{ii}^3 - A_{ii}^3 - (A^T)_{ii}^3}{2k_i^{\text{tot}}(k_i^{\text{tot}} - 1)}. \quad (\text{A8})$$

This expression counts a subset of the total number of triangles and is a special case of the expression derived in Ref. [115]. We will use this expression to measure the transitive clustering coefficient of each music networks (code available at Ref. [112]).

7. Spearman correlation coefficient and its associated p value

To assess the correlation between two quantities, we calculate the Spearman correlation coefficient between them. This coefficient is a nonparametric measure of the monotonicity of the relationship between two data sets. The Spearman correlation coefficient ranges from -1 to 1 , with 0 implying no correlation. The sign of the correlation coefficient indicates the direction of association between the two variables. For two data sets denoted by X and Y , it is computed as follows:

$$r_s = \frac{\text{cov}(R(X), R(Y))}{\sigma_{R(X)}\sigma_{R(Y)}}, \quad (\text{A9})$$

where $R(X)$ and $R(Y)$ denote the rank of the data points, $\text{cov}(R(X), R(Y))$ denotes the covariance of the rank variables, and $\sigma_{R(X)}$ and $\sigma_{R(Y)}$ are the standard deviations of the rank variables.

The reported p value indicates the probability that uncorrelated data sets have a Spearman correlation at least as extreme as the one computed. In instances where our sample size is not very large, we calculate the p value by performing a permutation test.

- [1] S. J. Mithen, *The Singing Neanderthals: The Origins of Music, Language, Mind and Body* (Harvard University Press, USA, 2005).
- [2] G. F. Welch, M. Biasutti, J. MacRitchie, G. E. McPherson, and E. Himonides, The impact of music on human development and well-being, *Front. Psychol.* **11**, 1246 (2020).

- [3] I. Cross, Music, cognition, culture, and evolution, *Ann. N.Y. Acad. Sci.* **930**, 28 (2001).
- [4] S. McClary, The impromptu that trod on a loaf: Or how music tells stories, *Narrative* **5**, 20 (1997).
- [5] D. Miell, R. A. MacDonald, and D. J. Hargreaves, *Musical Communication* (Oxford University Press, Oxford, 2005).

- [6] K. R. Scherer and E. Coutinho, How music creates emotion: A multifactorial process approach, *The Emotional Power of Music: Multidisciplinary Perspectives on Musical Arousal, Expression, and Social Control* (Oxford University Press, Oxford, 2013), pp. 121–145.
- [7] S. Koelsch, Brain correlates of music-evoked emotions, *Nat. Rev. Neurosci.* **15**, 170 (2014).
- [8] A. J. Blood and R. J. Zatorre, Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion, *Proc. Natl. Acad. Sci.* **98**, 11818 (2001).
- [9] D. Huron, *Sweet Anticipation: Music and the Psychology of Expectation* (The MIT Press, USA, 2006).
- [10] B. Tillmann, B. Poulin-Charronnat, and E. Bigand, The role of expectation in music: From the score to emotions and the brain, *WIREs Cognit. Sci.* **5**, 105 (2014).
- [11] M. T. Pearce and G. A. Wiggins, Auditory expectation: The information dynamics of music perception and cognition, *Topics Cognit. Sci.* **4**, 625 (2012).
- [12] V. N. Salimpoor, D. H. Zald, R. J. Zatorre, A. Dagher, and A. R. McIntosh, Predictions and the brain: How musical sounds become rewarding, *Trends Cognit. Sci.* **19**, 86 (2015).
- [13] M. A. Rohrmeier and S. Koelsch, Predictive information processing in music cognition. A critical review, *Int. J. Psychophysiol.* **83**, 164 (2012).
- [14] L. B. Meyer, *Emotion and Meaning in Music*, ACLS Humanities E-Book (University of Chicago Press, Chicago, USA, 1956).
- [15] P. Vuust and C. D. Frith, Anticipation is the key to understanding music and the effects of music on emotion, *Behav. Brain Sci.* **31**, 599 (2008).
- [16] H. Egermann, M. T. Pearce, G. A. Wiggins, and S. McAdams, Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music, *Cogn. Affect. Behav. Neurosci.* **13**, 533 (2013).
- [17] S. Ferretti, On the modeling of musical solos as complex networks, *Inform. Sci.* **375**, 271 (2017).
- [18] S. Ferretti, On the complex network structure of musical pieces: Analysis of some use cases from different music genres, *Multimedia Tools Appl.* **77**, 16003 (2018).
- [19] X. F. Liu, K. T. Chi, and M. Small, Complex network structure of musical compositions: Algorithmic generation of appealing music, *Physica A* **389**, 126 (2010).
- [20] M. B. Nardelli, *MUSICNTWRK: Data Tools for Music Theory, Analysis and Composition, Perception, Representations, Image, Sound, Music: 14th International Symposium, CMMR 2019, Marseille, France* (Springer International Publishing, Cham, 2021), Vol. 12631, pp. 190–215.
- [21] C. W. Lynn, L. Papadopoulos, A. E. Kahn, and D. S. Bassett, Human information processing in complex networks, *Nat. Phys.* **16**, 965 (2020).
- [22] C. E. Shannon, A mathematical theory of communication, *Bell Syst. Tech. J.* **27**, 379 (1948).
- [23] S. T. Piantadosi, H. Tily, and E. Gibson, Word lengths are optimized for efficient communication, *Proc. Natl. Acad. Sci.* **108**, 3526 (2011).
- [24] J. B. Plotkin and M. A. Nowak, Language evolution and information theory, *J. Theor. Biol.* **205**, 147 (2000).
- [25] J.-P. Eckmann, E. Moses, and D. Sergi, Entropy of dialogues creates coherent structures in e-mail traffic, *Proc. Natl. Acad. Sci.* **101**, 14333 (2004).
- [26] K. Zhao, M. Karsai, and G. Bianconi, Entropy of dynamical social networks, *PLoS One* **6**, e28116 (2011).
- [27] M. Rosvall, A. Trusina, P. Minnhagen, and K. Sneppen, Networks and cities: An information perspective, *Phys. Rev. Lett.* **94**, 028701 (2005).
- [28] J. E. Cohen, Information theory and music, *Behav. Sci.* **7**, 137 (1962).
- [29] L. Hiller and C. Bean, Information theory analyses of four sonata expositions, *J. Music Theory* **10**, 96 (1966).
- [30] F. Gomez, T. Lorimer, and R. Stoop, Complex networks of harmonic structure in classical music, in *Nonlinear Dynamics of Electronic Systems* (Springer International Publishing, Cham, 2014), pp. 262–269.
- [31] J. P. Boon, A. Noullez, and C. Mommen, Complex dynamics and musical structure, *Interface* **19**, 3 (1990).
- [32] L. Liu, J. Wei, H. Zhang, J. Xin, and J. Huang, A statistical physics view of pitch fluctuations in the classical music from Bach to Chopin: Evidence for scaling, *PLoS One* **8**, e58710 (2013).
- [33] D. Kahneman, S. P. Slovic, P. Slovic, and A. Tversky, *Judgment under Uncertainty: Heuristics and Biases* (Cambridge University Press, Cambridge, 1982).
- [34] P. Dayan, Improving generalization for temporal difference learning: The successor representation, *Neural Comput.* **5**, 613 (1993).
- [35] 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).
- [36] 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. Human Behav.* **1**, 680 (2017).
- [37] J. Sherwin and P. Sajda, Musical experts recruit action-related neural structures in harmonic anomaly detection: Evidence for embodied cognition in expertise, *Brain Cognit.* **83**, 190 (2013).
- [38] M. Tervaniemi, T. Tupala, and E. Brattico, Expertise in folk music alters the brain processing of Western harmony, *Ann. N.Y. Acad. Sci.* **1252**, 147 (2012).
- [39] P. Vuust, E. Brattico, M. Seppänen, R. Näätänen, and M. Tervaniemi, Practiced musical style shapes auditory skills, *Ann. N.Y. Acad. Sci.* **1252**, 139 (2012).
- [40] S. Jentschke and S. Koelsch, Musical training modulates the development of syntax processing in children, *NeuroImage* **47**, 735 (2009).
- [41] M. E. Curtis and J. J. Bharucha, Memory and musical expectation for tones in cultural context, *Music Percept.* **26**, 365 (2009).
- [42] S. M. Demorest and L. Osterhout, ERP responses to cross-cultural melodic expectancy violations, *Ann. N.Y. Acad. Sci.* **1252**, 152 (2012).
- [43] T. Eerola, T. Himberg, P. Toiviainen, and J. Louhivuori, Perceived complexity of western and African folk melodies by western and African listeners, *Psychol. Music* **34**, 337 (2006).
- [44] T. Eerola, J. Louhivuori, and E. Lebaka, Expectancy in Sami Yoiks revisited: The role of data-driven and schema-driven knowledge in the formation of melodic expectations, *Music. Sci.* **13**, 231 (2009).
- [45] T. Bent, A. R. Bradlow, and B. A. Wright, The influence of linguistic experience on the cognitive processing of pitch in

- speech and nonspeech sounds. *J. Exp. Psychol. Hum. Percept. Perform.* **32**, 97 (2006).
- [46] B. Chandrasekaran, A. Krishnan, and J. T. Gandour, Relative influence of musical and linguistic experience on early cortical processing of pitch contours, *Brain Lang.* **108**, 1 (2009).
- [47] P. Q. Pfordresher and S. Brown, Enhanced production and perception of musical pitch in tone language speakers, *Atten. Percept. Psychophys.* **71**, 1385 (2009).
- [48] M. P. Roncaglia-Denissen, D. A. Roor, A. Chen, and M. Sadakata, The enhanced musical rhythmic perception in second language learners, *Frontiers Human Neurosci.* **10**, 288 (2016).
- [49] M. W. Howard and M. J. Kahana, A distributed representation of temporal context, *J. Math. Psychol.* **46**, 269 (2002).
- [50] S. J. Gershman, C. D. Moore, M. T. Todd, K. A. Norman, and P. B. Sederberg, The successor representation and temporal context, *Neural Comput.* **24**, 1553 (2012).
- [51] C. W. Lynn and D. S. Bassett, How humans learn and represent networks, *Proc. Natl. Acad. Sci.* **117**, 29407 (2020).
- [52] G. R. Jagari, P. Pedram, and L. Hedayatifar, Long-range correlation and multifractality in Bach's invention pitches, *J. Stat. Mech.: Theory Exp.* (2007) P04012.
- [53] J. M. Moore, D. C. Corrêa, and M. Small, Is Bach's brain a Markov chain? Recurrence quantification to assess Markov order for short, symbolic, musical compositions, *Chaos* **28**, 085715 (2018).
- [54] F. Lerdahl and R. S. Jackendoff, *A Generative Theory of Tonal Music, Reissue, with a New Preface* (MIT Press, USA, 1996).
- [55] S. Koelsch, M. Rohrmeier, R. Torrecuso, and S. Jentschke, Processing of hierarchical syntactic structure in music, *Proc. Natl. Acad. Sci.* **110**, 15443 (2013).
- [56] J. Gomez-Gardenes and V. Latora, Entropy rate of diffusion processes on complex networks, *Phys. Rev. E* **78**, 065102(R) (2008).
- [57] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevResearch.6.013136> for more information.
- [58] R. Albert, Scale-free networks in cell biology, *J. Cell Sci.* **118**, 4947 (2005).
- [59] K. Lerman, X. Yan, and X.-Z. Wu, The majority illusion in social networks, *PLoS One* **11**, e0147617 (2016).
- [60] A.-L. Barabási and R. Albert, Emergence of scaling in random networks, *Science* **286**, 509 (1999).
- [61] M. M. Garvert, R. J. Dolan, and T. E. J. Behrens, A map of abstract relational knowledge in the human hippocampal-entorhinal cortex, *elife* **6**, e17086 (2017).
- [62] F. Meyniel, M. Maheu, and S. Dehaene, Human inferences about sequences: A minimal transition probability model, *PLoS Comput. Biol.* **12**, e1005260 (2016).
- [63] F. Meyniel and S. Dehaene, Brain networks for confidence weighting and hierarchical inference during probabilistic learning, *Proc. Natl. Acad. Sci.* **114**, E3859 (2017).
- [64] E. L. Newport and R. N. Aslin, Learning at a distance I. Statistical learning of non-adjacent dependencies, *Cognit. Psychol.* **48**, 127 (2004).
- [65] A. E. Kahn, E. A. Karuza, J. M. Vettel, and D. S. Bassett, Network constraints on learnability of probabilistic motor sequences, *Nat. Hum. Behav.* **2**, 936 (2018).
- [66] J. R. Saffran, E. K. Johnson, R. N. Aslin, and E. L. Newport, Statistical learning of tone sequences by human infants and adults, *Cognition* **70**, 27 (1999).
- [67] E. Morgan, A. Fogel, A. Nair, and A. D. Patel, Statistical learning and gestalt-like principles predict melodic expectations, *Cognition* **189**, 23 (2019).
- [68] V. K. Cheung, P. M. C. Harrison, S. Koelsch, M. T. Pearce, A. D. Friederici, and L. Meyer, Cognitive and sensory expectations independently shape musical expectancy and pleasure, *Phil. Trans. R. Soc. B* **379**, 20220420 (2023).
- [69] T. Collins, B. Tillmann, F. S. Barrett, C. Delbé, and P. Janata, A combined model of sensory and cognitive representations underlying tonal expectations in music: From audio signals to behavior, *Psychol. Rev.* **121**, 33 (2014).
- [70] M. T. Pearce, M. H. Ruiz, S. Kapasi, G. A. Wiggins, and J. Bhattacharya, Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation, *NeuroImage* **50**, 302 (2010).
- [71] D. J. Watts and S. H. Strogatz, Collective dynamics of 'small-world' networks, *Nature (London)* **393**, 440 (1998).
- [72] G. Szabó, M. J. Alava, and J. Kertész, Clustering in complex networks, *Lect. Notes Phys.* **650**, 139 (2004).
- [73] I. Mencke, D. Omigie, M. Wald-Fuhrmann, and E. Brattico, Atonal music: Can uncertainty lead to pleasure? *Front. Neurosci.* **12**, 979 (2019).
- [74] B. Tillmann and E. Bigand, The relative importance of local and global structures in music perception, *J. Aesth. Art Criticism* **62**, 211 (2004).
- [75] D. Butler, Describing the perception of tonality in music: A critique of the tonal hierarchy theory and a proposal for a theory of intervallic rivalry, *Music Percept.* **6**, 219 (1989).
- [76] P. Toiviainen and C. L. Krumhansl, Measuring and modeling real-time responses to music: The dynamics of tonality induction, *Perception* **32**, 741 (2003).
- [77] J. R. Saffran, R. N. Aslin, and E. L. Newport, Statistical learning by 8-month-old infants, *Science* **274**, 1926 (1996).
- [78] J. Fiser and R. N. Aslin, Statistical learning of higher-order temporal structure from visual shape sequences. *J. Exp. Psychol.* **28**, 458 (2002).
- [79] A. R. Romberg and J. R. Saffran, Statistical learning and language acquisition, *Wiley Interdiscip. Rev.: Cognit. Sci.* **1**, 906 (2010).
- [80] R. F. I. Cancho and V. S. Richard, The small world of human language, *Proc. R. Soc. London B* **268**, 2261 (2001).
- [81] P. H. R. Zivic, F. Shifres, and G. A. Cecchi, Perceptual basis of evolving Western musical styles, *Proc. Natl. Acad. Sci.* **110**, 10034 (2013).
- [82] C. Pérez-Sancho, D. Rizo, and J. M. Iñesta, Genre classification using chords and stochastic language models, *Connect. Sci.* **21**, 145 (2009).
- [83] D. K. Simonton, Melodic structure and note transition probabilities: A content analysis of 15,618 classical themes, *Psychol. Music* **12**, 3 (1984).
- [84] J. Thickstun, Z. Harchaoui, and S. Kakade, Learning features of music from scratch, [arXiv:1611.09827](https://arxiv.org/abs/1611.09827).
- [85] I. Liu, B. Ramakrishnan *et al.*, Bach in 2014: Music composition with recurrent neural network, [arXiv:1412.3191](https://arxiv.org/abs/1412.3191).

- [86] D. Collins, A synthesis process model of creative thinking in music composition, *Psychol. Music* **33**, 193 (2005).
- [87] D. Collins, Real-time tracking of the creative music composition process, *Digital Creativity* **18**, 239 (2007).
- [88] W. J. Dowling and J. C. Bartlett, The importance of interval information in long-term memory for melodies, *Psychomusicology* **1**, 30 (1981).
- [89] R. Parncutt, *Harmony: A Psychoacoustical Approach*, Springer Science & Business Media, Vol. 19 (Springer, Berlin, Heidelberg, 2012).
- [90] C. J. Stevens, Music perception and cognition: A review of recent cross-cultural research, *Topics Cognit. Sci.* **4**, 653 (2012).
- [91] F. Lerdahl and R. Jackendoff, An overview of hierarchical structure in music, *Music Percept.* **1**, 229 (1983).
- [92] J. Xu, T. L. Wickramaratne, and N. V. Chawla, Representing higher-order dependencies in networks, *Sci. Adv.* **2**, e1600028 (2016).
- [93] R. Lambiotte, M. Rosvall, and I. Scholtes, From networks to optimal higher-order models of complex systems, *Nat. Phys.* **15**, 313 (2019).
- [94] H. Yin, A. R. Benson, and J. Leskovec, Higher-order clustering in networks, *Phys. Rev. E* **97**, 052306 (2018).
- [95] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, Multilayer networks, *J. Compl. Netw.* **2**, 203 (2014).
- [96] A. Aleta and Y. Moreno, Multilayer networks in a nutshell, *Annu. Rev. Condens. Matter Phys.* **10**, 45 (2019).
- [97] M. De Domenico, A. Solé-Ribalta, E. Cozzo, M. Kivelä, Y. Moreno, M. A. Porter, S. Gómez, and A. Arenas, Mathematical formulation of multilayer networks, *Phys. Rev. X* **3**, 041022 (2013).
- [98] Stefano Boccaletti, Ginestra Bianconi, R. Criado, C. I. Del Genio, J. Gómez-Gardenes, M. Romance, I. Sendina-Nadal, Z. Wang, and M. Zanin, The structure and dynamics of multilayer networks, *Phys. Rep.* **544**, 1 (2014).
- [99] M. De Domenico, C. Granell, M. A. Porter, and A. Arenas, The physics of spreading processes in multilayer networks, *Nat. Phys.* **12**, 901 (2016).
- [100] J. R. Iversen, A. D. Patel, and K. Ohgushi, Perception of rhythmic grouping depends on auditory experience, *J. Acoust. Soc. Am.* **124**, 2263 (2008).
- [101] J. Stiller, D. Nettle, and R. I. Dunbar, The small world of Shakespeare's plays, *Human Nature* **14**, 397 (2003).
- [102] Y.-M. Choi and H.-J. Kim, A directed network of greek and roman mythology, *Physica A* **382**, 665 (2007).
- [103] M. J. Buehler, Unsupervised cross-domain translation via deep learning and adversarial attention neural networks and application to music-inspired protein designs, *Patterns* **4**, 100692 (2023).
- [104] C.-H. Yu, Z. Qin, F. J. Martin-Martinez, and M. J. Buehler, A self-consistent sonification method to translate amino acid sequences into musical compositions and application in protein design using artificial intelligence, *ACS Nano* **13**, 7471 (2019).
- [105] Joyce Y. Wong, John McDonald, Micki Taylor-Pinney, David I. Spivak, David L. Kaplan, and Markus J. Buehler, Materials by design: Merging proteins and music, *Nano Today* **7**, 488 (2012).
- [106] https://github.com/SumanSKulkarni/Music_Networks
- [107] Bach central, <http://www.bachcentral.com/midiindexcomplete.html>.
- [108] Kern scores, <http://kern.humdrum.org/search?s=t&keyword=Bach+Johann>.
- [109] Bach cantatas website, <https://www.bach-cantatas.com/Mus/BWV1-Mus.htm>.
- [110] Suzu MIDI, <http://www.suzumidi.com/eng/bach4.htm>.
- [111] MIDI file tools for MATLAB by Ken Schutte, <https://github.com/kts/matlab-midi>.
- [112] Information content of note transitions in the music of J. S. Bach. https://github.com/SumanSKulkarni/Music_Networks.
- [113] M. E. J. Newman and M. Girvan, Finding and evaluating community structure in networks, *Phys. Rev. E* **69**, 026113 (2004).
- [114] F. Váša and B. Mišić, Null models in network neuroscience, *Nat. Rev. Neurosci.* **23**, 493 (2022).
- [115] G. Fagiolo, Clustering in complex directed networks, *Phys. Rev. E* **76**, 026107 (2007).