Eur. Phys. J. B (2022)95:103
https://doi.org/10.1140/epjb/s10051-022-00368-z

THE EUROPEAN PHYSICAL JOURNAL B

Regular Article - Statistical and Nonlinear Physics

Harmonic structures of Beethoven quartets: a complex network approach

Theo Frottier[1,a], Bertrand Georgeot[1,b], and Olivier Giraud[2,c]

[1] Laboratoire de Physique Théorique, Université de Toulouse,CNRS, UPS, 31062 Toulouse, France
[2] Université Paris Saclay, CNRS, LPTMS, 91405 Orsay, France

Received 7 April 2022 / Accepted 15 June 2022

Abstract. We propose a complex network approach to the harmonic structure underpinning western tonal music. From a database of Beethoven's string quartets, we construct a directed network whose nodes are musical chords and edges connect chords following each other. We show that the network is scale-free and has specific properties when ranking algorithms are applied. We explore the community structure and its musical interpretation, and propose statistical measures stemming from network theory allowing to distinguish stylistically between periods of composition. Our work opens the way to a network approach of structural properties of tonal harmony.

1 Introduction

In the recent past, network theory has been developed as a new tool enabling to uncover structural properties, dynamics and evolution of a variety of systems, from natural ones, such as biological networks, to human produced ones such as the World Wide Web or social networks [1]. Interestingly enough, it has also been shown that this theory can give new insights on systems which are not obviously organized as networks, such as languages [2–5] or board games [6–9].

Music shares features from both languages and games. Connections between music and natural sciences are numerous, be it physiology, physics of waves, or group theory [10]. The improvement of computer capabilities has recently opened several lines of research at the interface between musicology, mathematics and computer science, from automatic harmonic analysis [11,12], statistical analysis of music [13], creation of databases [14,15], computer-assisted Schenkerian analysis [16], to recent applications from machine learning [17,18]. A huge corpus of musical pieces exists with many musicological studies analyzing their history and evolution (see e.g. [19,20] for a historical account). Importantly, musical syntax is not so much about the perception of isolated chords as about the relationship between a chord and the ones that surround it. An important aspect of musical analysis is thus to understand how chords are interrelated, both at a global and at a local scale, the latter corresponding to the neighbourhood of a chord in a given musical segment. In this paper, our aim is to apply network theory to musical pieces.

In 2018, a database of all chords of Beethoven string quartets was established [14], based on harmonic analyses made by human experts from the musical scores of the quartets. This database was analysed in [21], where the authors investigated the frequency distribution of chords, pairs of chords, and higher-order n-grams.

In the present paper we go a step further by constructing a network based on temporal relations between chords within musical segments, following ideas borrowed from text analysis [2–5]. In the following, we build the network, discuss its properties and their relationship with musical features, and investigate its community structure. The string quartets are particularly interesting in that their composition stretches over a period of 28 years of Beethoven's life, allowing to follow the stylistic evolution throughout his lifetime [22–24]. Here we show that tools from network theory allow to statistically differentiate between the different periods of the Beethoven quartets.

The present network approach bears some analogy with the Euler Tonnetz, the geometry of musical chords [25], or a network approach of atonal music in [26]. A network of chord progressions, similar in spirit to ours, was proposed very recently in [27], based on chords taken as vertical arrangements of pitch classes and a small-scale analysis of data. By contrast, our approach considers chords in a functional relationship with a local key, as determined by human experts, and over the scale of a whole corpus. Our work shows that a network approach provides some insight into structural properties of tonal harmony.

[a] e-mail: theo.frottier@insa-toulouse.fr
[b] e-mail: georgeot@irsamc.ups-tlse.fr
[c] e-mail: olivier.giraud@universite-paris-saclay.fr (corresponding author)

2 The database

The annotated database of all chords from the complete set of Beethoven string quartets (16 quartets, a total of 70 movements) is available online at [28]. It has 28,095 entries, each of which provides information on a chord : the global and local key with respect to which it appears, as well as possible changes in the chord, relative root, or pedals. It also contains information on the chord duration, movement and measure in which it appears, and whether or not it is at the end of a musical segment.

Each chord is characterized by a Latin numeral (from I to VII) indicating its relation with the local key, and a figure (6, 64, 7, 65, 43, 2) indicating whether the chord is in root position or appears as an inversion. It may also contain an indication of its form (major seventh, half-diminished, diminished and augmented, respectively denoted M, %, $\circ$, and $+$), various figures between brackets indicating changes in the inversion, and/or additional Latin numbers indicating the relative root. For some of the chords (107 of them in the database), no harmonic value could be determined by the experts, and a label "none" was assigned; such chords are most frequent in the late quartets (with 58 undetermined chords), which are known to be more innovative.

The whole corpus is divided into 929 musical segments. Each has a specific local key, which indicates whether the segment is major or minor. We split the database into two parts, one for chords in segments with major local key and one for chords in minor local key. Among the 929 musical segments there are 551 major and 378 minor ones, yielding a database of 20,276 chords in major segments and 7819 chords in minor segments. Among these 20,276 entries in major segments we found $N_{\mathcal{M}} = 871$ distinct ones. The minor segments involve $N_{\mathrm{m}} = 599$ distinct entries.

Individual chords can be ranked by their number of occurrences in the database. We found that the frequency f of occurences as a function of the rank r has a power–law tail $f \propto 1/r^{\gamma}$. This is characteristic of the Zipf law, which was first observed in the analysis of languages [29] and since then in many contexts. For our database, a similar power–law tail was already obtained in [21]. Here we find an exponent $\gamma \approx 1.6$, not so far from the exponent ≈ 2 found in natural languages.

3 Network theory

To go beyond mere statistics of chords, we now introduce the basic tools of network theory.

3.1 The networks

A graph is a set of vertices connected by edges. In our case, we construct a graph $\mathcal{M}$ based on chords from major segments. Its vertices are the $N_{\mathcal{M}}$ distinct chords

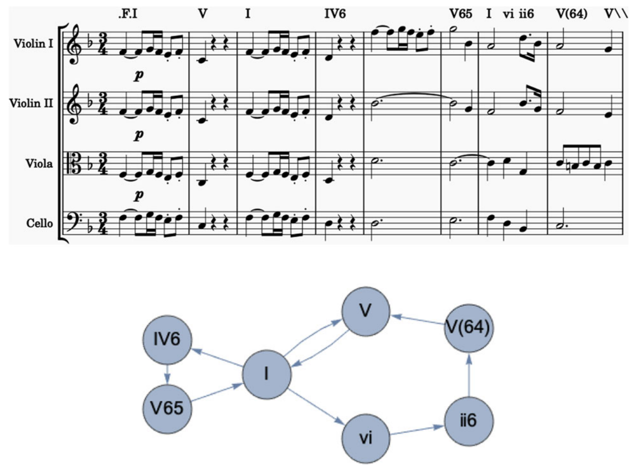

Fig. 1 First segment of Beethoven's Op. 18 No. 1 in F major (first movement) and its associated graph. The score was generated from the data of [28] using MuseScore [30]

appearing in the database. We also construct a graph $\mathfrak{m}$ with chords from minor segments, yielding a graph with $N_{\mathfrak{m}}$ vertices.

Our networks are built in the following way: We add a directed edge between two vertices i and j each time chord j immediately follows chord i in the same segment. There are as many edges between i and j as occurrences of the pair i, j in the database, which makes our graph a weighted directed graph. As an illustration, in Fig. 1 we show the first 8 bars of Beethoven's Op. 18 No. 1, which correspond to the first segment of the database. It consists of 10 labeled chords (7 distinct ones). The corresponding graph, with 7 vertices and 9 edges, is given below the score.

3.2 The PageRank algorithm

One of the tools developed for investigating the network structure is the PageRank algorithm, which gave the original impulse to the development of the Google search engine [31]. The PageRank algorithm is built to hierarchize the nodes of a network in a relevant way, by constructing a vector (the PageRank vector) whose entries are used to rank the vertices by order of importance. This vector is the eigenvector associated with the largest eigenvalue of a matrix G constructed from the $N \times N$ adjacency matrix of the network.

This Google matrix G is defined, for some parameter α in $[0, 1]$, as $G_{ij} = \alpha S_{ij} + (1 - \alpha)/N$, where S is obtained from the weighted adjacency matrix by replacing any column containing only 0 by a column of $1/N$ and normalizing the sum of entries of each column to 1. In the case of the quartet database, we did not encounter any such column of zeros, as segments end up with chords which are frequent in the database. Adding the constant part proportional to $1 - \alpha$ to that matrix S avoids numerical results being dominated by dangling groups (that is, groups of vertices with no outgoing edges), as those tend to dominate the PageRank when $\alpha \to 1$.

By construction, G is a stochastic matrix. Perron–Frobenius theorem ensures that G has an eigenvector with eigenvalue 1 and real positive entries. The PageRank vector p is defined as the vector such that $Gp = p$ and $\sum_i p_i = 1$. The value p_i can be interpreted as the probability for a random surfer following the edges of the network for an infinite time to be found on vertex i, if at each step the outgoing edges are chosen at random with equal probability. Properties of the Google matrix shed light on the structure of the network; this approach was successfully applied in a number of contexts [32].

3.3 Network communities

The most basic structure that underlies the topology of a graph is its partition into communities, that is, subsets that have more connections within themselves than between one another. A way of determining whether a given partition of the set of vertices properly describes its community structure is to compute the modularity of that partition. This quantity measures how far a given graph is from a graph with the same connectivity but with edges taken at random within and between subsets of the partition. For a given partition into communities $\{\mathcal{C}\}$, the modularity is given by $\mu = \sum_{\mathcal{C}} \sum_{i,j\in\mathcal{C}} [a_{ij} - d_i d_j/(2m)]$, where a_{ij} is equal to the number of undirected edges connecting vertices i and j, d_i is the total number of edges from i, m is the total number of edges in the undirected graph, and the sum runs over all communities $\mathcal{C}$. The partition that yields the highest modularity provides a possible decomposition of the graph into communities [33,34].

4 The musical network

We now apply the above tools to our musical networks. In Fig. 2 (top left panel) we display the cumulative distribution of incoming and outgoing edges, that is, the number of vertices that have more than k ingoing (or outgoing) edges, with k normalized by its maximum value. It follows a power–law $P_{\text{in,out}} \sim 1/k^{\nu}$ with exponent $\nu \approx 0.7$. Similar power–law distributions of vertex connectivities were found in many real-world complex networks, known as scale-free networks [35,36]. Here the exponent $1+\nu$ of the (non-integrated) distribution roughly corresponds to the exponent $\gamma = 1.6$ found for the Zipf law. Typically in scale-free networks the exponent $1+\nu$ ranges from 2 to 3, but lower exponents have been found, for example ≈ 1.5 for e-mail networks [37].

4.1 PageRank vector

In the top right panel of Fig. 2 we show the ranking of chords as given by the PageRank vector. As shown in the lower panels of Fig. 2, this ranking is quite different from that given by the mere frequency of chords in the database. Some chords, although rare in the database, have a high PageRank; they can correspond to rare followers of much more common patterns. This is the case, for instance, for the chord labeled bIII, which generally follows the high-rank chord I. The PageRank vector follows a power–law $p_i \sim 1/i^{\beta}$ with $\beta \approx 0.93$, very close to the exponent 0.9 found for networks describing parts of the World Wide Web [38–41].

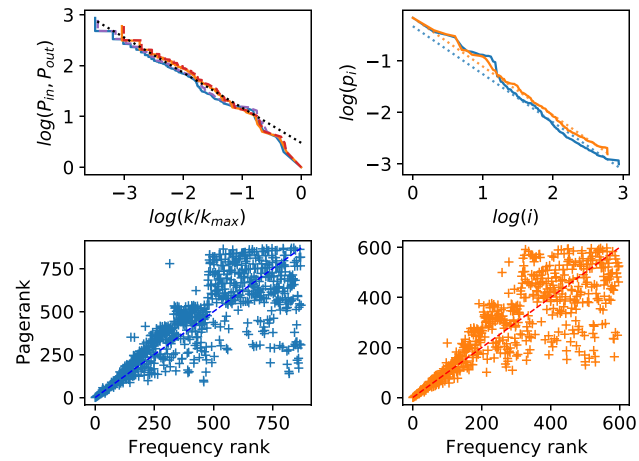

Fig. 2 Top left: Cumulative distribution of ingoing edges P_{in} (major blue, minor orange) and outgoing edges P_{out} (major violet, minor red, almost indistinguishable from P_{in}). A linear fit over the 30 leftmost points yields slopes −0.68 for major $P_{\text{in,out}}$ and −0.70 for minor $P_{\text{in,out}}$. Dotted black line has a slope −0.69. Top right: PageRank vector p_i (sorted in descending order) for graphs $\mathcal{M}$ (blue) and $\mathfrak{m}$ (orange). Dotted lines are linear fit, with slope −0.93 (major) and −0.94 (minor). Bottom: PageRank vs frequency rank for the major (left, blue) and minor (right, orange). Each point represents a specific chord. Dashed line is the line $y = x$

4.2 Spectrum of the Google matrix

The spectrum of G gives some insight into the structure of the network. For a symmetric matrix the spectrum is real. For directed networks the matrix G is in general non-symmetric, and the complex spectrum is all the more flattened onto the real axis as there exist pairs of edges of opposite directions between pairs of vertices. For example, this happens for dictionary networks, where many words are symmetrically related [41]. The spectrum of G, displayed in Fig. 3, shows that there is no such phenomenon, consistently with the temporal directionality of music. For instance, in major segments there are in total 40 occurences of the pair ii→V, but only 7 of the pair V→ii. This aspect of music, referred to as directedness, is also discussed in [21].

From Perron–Frobenius theorem, the spectrum of G is by construction bounded by the circle of radius α, except for the lone eigenvalue corresponding to the PageRank. However, it is clear from Fig. 3 that the spectrum is concentrated much closer to the center than the theoretical bound, almost entirely within a circle

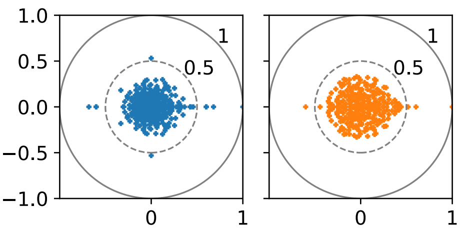

Fig. 3 Spectrum of G in the complex plane (left major, right minor) for $\alpha = 0.85$. Dashed circle of radius $\frac{1}{2}$ is an eyeguide

of radius $\frac{1}{2}$. This reflects the fact that the network is highly connected, with many edges between different parts. Indeed, eigenvalues with large modulus correspond to long-lived eigenstates located on parts of the network which are less connected with the rest. For a dense graph, isolated regions are rarer, which tends to suppress such outlying eigenvalues. In the present case, the smaller radius of the spectrum can be interpreted as a reflection of the fact that the same chords can appear in many different contexts, which homogeneizes the graph. The spectrum of $\mathfrak{m}$ is less concentrated, indicating that this phenomenon is less pronounced in minor. Lastly, we can notice in Fig. 3 the presence of isolated eigenvalues separated from the main cluster of eigenvalues. As will be illustrated below on the spectra of Fig. 5, this reflects zones of the graph where groups of nodes are weakly coupled to other parts (for example, patterns of chords that appear only in specific contexts). Such features correspond to the notion of communities in a graph, to which we now turn.

4.3 Communities

As mentioned earlier, communities can be obtained by computing the modularity of partitions of the undirected graph and identifying the partition of maximal modularity. There is a variety of ways of computing the partition with highest modularity. We use the Louvain algorithm [42], implemented in the NetworkX package of Python. The output of the algorithm depends on a random seed, and for a given graph the resulting maximal modularity changes (mildly) from one run of the program to the other, as well as the partition itself. Nevertheless, the main features of the communities are robust.

In Fig. 4 we show the community partition for our graphs. We illustrate our results with the outcome with the highest modularity, namely 0.2252 for the graph $\mathcal{M}$ and 0.2572 for the graph $\mathfrak{m}$. For the graph $\mathcal{M}$, we find 5 main communities, each of which revolves around an elementary chord: I, IV, V, I6 and ii6. Within these communities one finds closely related chords. For instance, the community "IV" localizes on the subdominant IV but also on chords which have a close harmonic function with respect to the subdominant (such as V2/IV, V7/IV). The small outlying communities, very weakly connected with the rest of the graph, correspond to sequences of rare chords (such as chords appearing only once in the corpus).

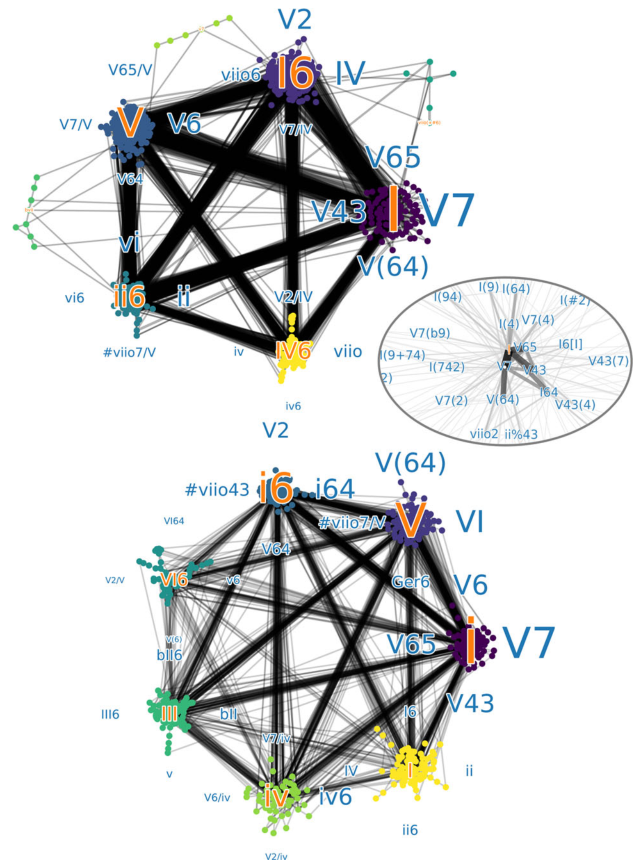

Fig. 4 Community partition of the networks (top major, bottom minor). At the centre of each cloud (in orange) is the chord with largest p_i (highest PageRank) in the community; the next four chords in PageRank order are labeled in blue around each community. Symbol size reflects value of p_i. Inset: zoom on the community I of $\mathcal{M}$

This community structure reflects the presence of poles of attraction, a dimension of music referred to as centricity in [21]. As can be expected from a musical perspective, these poles include (in major) the fourth degree IV, dominant V, and tonic I. But interestingly, inversions of V belong to the same community as I. Moreover the other poles of attraction are also surrounded by inversions of their relative dominant, which shows that they locally behave as the tonic. Similar features are found in the graph $\mathfrak{m}$ in Fig. 4 bottom. The partition into communities thus yields a mesoscopic picture and allows to assess the role of chords within a community.

5 Comparison between the different periods

We now analyze how network properties depend on the period of composition of the quartets. It is well-known that Beethoven underwent a strong stylistic evo-

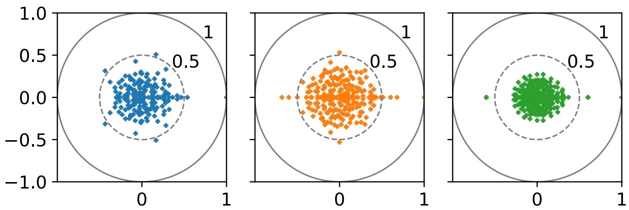

Fig. 5 From left to right: spectrum of G in the complex plane for the early, middle and late periods for the major dataset

lution during his lifetime, the last quartets in particular being more akin to romantic music that the previous ones. It is usually recognized that three temporally well-separated periods can be identified. The early period corresponds to the first six quartets (1798–1800), the middle period to the five next (1806–1814), and the late quartets are the last five (1824–1826), the Große Fuge being the last movement of Op. 130. In order to assess how this evolution reflects in the musical networks, for each period (early/middle/late) and each mode (major/minor) we constructed a network by only taking into account the corresponding musical segments. For major segments, the number of chords is 6496 for the early period, 4794 for the middle period, and 8986 for the late period. This results in graphs of size 379, 442 and 574, corresponding to the respective number of distinct chords.

5.1 Spectrum

In Fig. 5 we display the spectra of G associated with each period for major segments. As in Fig. 3 for the global network, the spectrum is concentrated near the origin, all the more so for the last period. As this concentration is linked with a strong connectivity between different parts of the network, this latter feature can be associated with the fact that the last period of Beethoven's quartets is associated with more musical originalities and formal innovation, leading to the appearance of new edges between previously unconnected chords. Some isolated eigenvalues can be found outside the region where most eigenvalues concentrate. In particular one can observe the presence of families of eigenvalues located at the same radius. For instance, near the dashed circle in Fig. 5 left are 5 eigenvalues separated by $2\pi/5$; these eigenvalues are associated with closely related eigenvectors, which are essentially located on the same chords (here, by order of amplitude, bIII, V65/bIII, iv, V6/bIII). They correspond to groups of chords usually played together, and are relatively well separated from the rest of the network. These groups play the role of small communities. This type of eigenstates is clearly less prominent in the third network, indicating again that the last period of Beethoven's quartets contains more musical innovations, connecting isolated groups of chords in the network.

5.2 Communities

The partition of the network into communities also presents different characteristics from one period to another. The community partition yielding the highest modularity over different trials of the Louvain algorithm corresponds to modularities 0.2651, 0.2607, and 0.2290, respectively. This is to be compared with the (maximal) modularity for graphs with same vertex degree distribution but random edges: for the corresponding sizes, modularity is respectively 0.091 ± 0.0034 for $N = 379$, 0.126 ± 0.0036 for $N = 442$, and 0.094 ± 0.0028 for $N = 574$ (the standard deviation is obtained from 1000 realizations). The decrease of modularity from early to late is therefore significant. As already pointed out above for the spectrum, this result is in line with the greater homogeneity of the late graph, due to the presence of new edges between chords not previously connected.

The community partition for the three graphs in major is displayed in Fig. 6. The weaker modularity of the late period is a manifestation of the larger connectivity between communities.

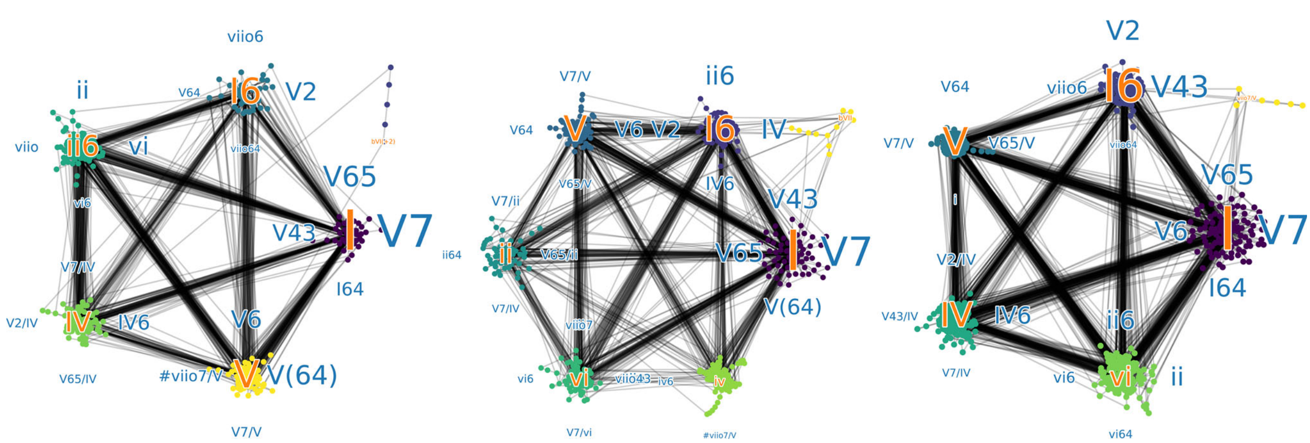

Fig. 6 Left to right: community structure of early, middle and late graphs of the major dataset, same conventions as in Fig. 4

5.3 Statistical analysis of the evolution of Beethoven's quartets

We now assess whether the stylistic evolution of the quartets also reflects in statistically significant differences in the networks, using the PageRank vectors. The fidelity is defined as the square of the scalar product between the two vectors (normalized in such a way that $\sum_i |\psi_i|^2 = 1$), namely $F(\psi, \phi) = |\sum_i \psi_i^* \phi_i|^2$. In Fig. 7 we display fidelities between different PageRanks. In order to assess their statistical significance, we construct several networks for each period: namely, for any pair of quartets a, b of a given period and mode, we construct the graph $\mathcal{G}_{ab}$ based on chords from quartets a and b only, and the corresponding PageRank vector p. For early major, which has 6 quartets, we get 15 different PageRank vectors, while middle and late major (with 5 quartets) have 10 PageRank vectors each. The points in Fig. 7 top correspond to all possible fidelities between these vectors, with the restriction that pairs of vectors from the same period should not contain any quartet in common (for instance there are 45 such pairs when comparing the early period with itself, and 150 when comparing the early period with the middle one). The vertical line and parses [·|·] indicate the mean and standard deviation of the values. As appears in the plot, the mean values of fidelities within a period (first three lines) are centered around 0.955. By contrast, the fidelities of pageranks from different periods are statistically weaker. The largest difference is between periods 1 and 3.

In order to compare more precisely the rankings between the different periods, we introduce the 'similarity' S between vectors, which we define as the mean number of identical indices within the $m = 30$ first indices when they are ordered by decreasing p_i. The results are plotted at Fig. 7 bottom. Although the results are less statistically significant than for the fidelity, there is a clear difference between the late period and the first two, and the first and second period are much more similar.

Our results are thus compatible with the opinion of musical scholars, showing a marked difference between the late period and the early and middle ones. This indicates that such tools from network theory are able to capture stylistic differences in music.

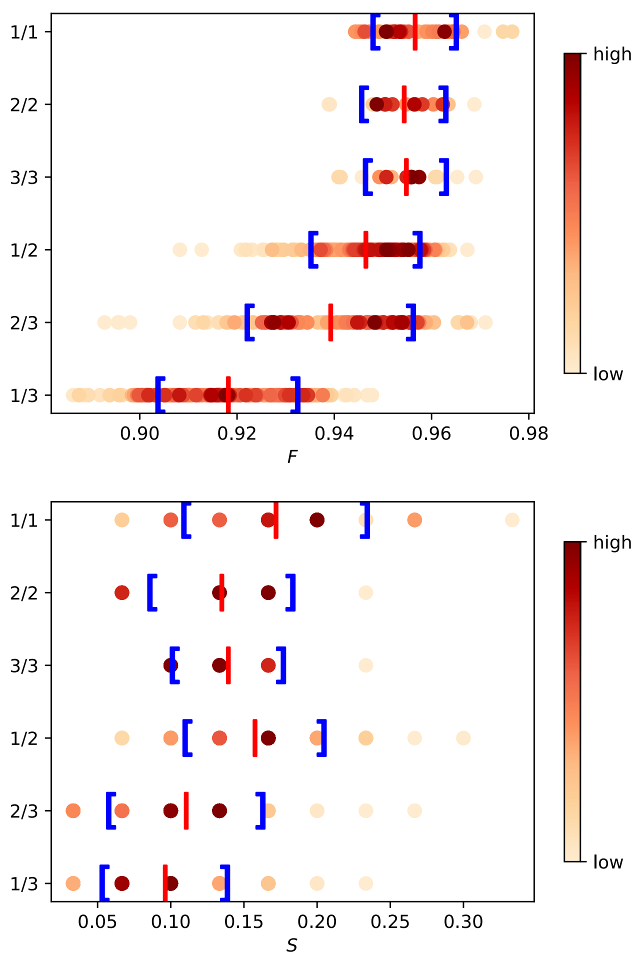

Fig. 7 Top: distribution of the fidelity F for different periods. Each point at line i/j corresponds to the fidelity between a pair of distinct PageRank vectors from periods i and j (see text for detail). The color represents the density of points, calculated by counting the number of values within a distance of 0.002 from the actual point. Bottom: same for similarity S (note that the discreteness of S makes the values highly degenerate)

6 Conclusion

The present work shows that the complex network approach can be fruitfully applied to the harmonic structure of musical works. Based on the example of Beethoven string quartets, we specified the properties of this new type of networks, and in particular we discussed the relationship between the spectrum of the Google matrix, the community structures, and musical specificities of the scores. We have also shown that the tools of complex networks allow to distinguish between the different periods and styles of Beethoven string quartets.

Our work opens the way to similar statistical analyses for different composers. In 1815 the Allgemeine musikalische Zeitung wrote that "Beethoven is without question the boldest sailor on the tide of harmony" (cited by [22]); similar harmonic analyses of pieces by Palestrina [43], Bach [13], Mozart [15] or Schubert [44] from a network approach would thus give an interesting perspective. Other aspects of music, such as history of harmonic patterns [45], or rhythm [46], could benefit from the network approach. Another possible fruitful direction could be to apply this approach to uncover hierarchical structures in music, in the spirit of Schenkerian analysis [47], by performing some coarse-graining to the network.

Acknowledgments We thank A. Ouali, who was involved in a preliminary study. We thank Calcul en Midi-Pyrénées (CalMiP) for access to its supercomputers.

Data Availibility Statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.]

Appendix A Cleaning the database

We found several issues in the database, such as missing data or faulty labels, most of which are listed at the database website [28]. We made the following corrections to the file all_annotations.tsv available at [28]. Global keys labeled 'nothing' or 'false' have been replaced by their correct value, given in the first column of the database. For some entries, the local key was labeled 'Ab' instead of 'VI': they were restored to their correct value. The local keys labeled 'I' at the beginning of some minor segments were relabeled 'i'.

We also chose to consider chords within a pedal segment to be treated without reference to the pedal (although the first chord of a pedal segment is treated as distinct). As for entries labeled 'none', i.e. chords for which no consensual harmonic analysis could be extracted from the score, we chose to treat them as a chord on its own.

In order to check the consistency of the corrections we made to the database, we compared our results with the ones obtained in [21]. In particular, for both major and minor segments, we calculated the list of frequencies of each chord type and the heatmaps (frequency of each sequence of pairs of chords), following [21]; the numerical outcomes we obtain is close to the ones obtained in [21]. The main difference is the frequency of 'I' in minor segments, which ranks 14 in frequency order in our database but 2 in Ref. [21]. It is very likely that the corrections listed in [28] have been performed after [21] was published, which would explain this discrepancy.

References

1. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Complex networks: structure and dynamics. Phys. Rep. **424**(4–5), 175–308 (2006)
2. R.F.I. Cancho, R.V. Solé, The small world of human language. Proc. R. Soc. Lond. B **268**(1482), 2261–2265 (2001)
3. A.P. Masucci, G.J. Rodgers, Network properties of written human language. Phys. Rev. E **74**(2), 026102 (2006)
4. L. Antiqueira, Md.G.V. Nunes, O. Oliveira Jr, Ld.F. Costa, Strong correlations between text quality and complex networks features. Phys. A **373**, 811–820 (2007)
5. R.V. Solé, B. Corominas-Murtra, S. Valverde, L. Steels, Language networks: their structure, function, and evolution. Complexity **15**(6), 20–26 (2010)
6. B. Georgeot, O. Giraud, The game of go as a complex network. Europhys. Lett. **97**(6), 68002 (2012)
7. V. Kandiah, B. Georgeot, O. Giraud, Move ordering and communities in complex networks describing the game of go. Eur. Phys. J. B **87**(10), 1–13 (2014)
8. L.-G. Xu, M.-X. Li, W.-X. Zhou, Weiqi games as a tree: Zipf's law of openings and beyond. Europhys. Lett. **110**(5), 58004 (2015)
9. C. Coquidé, B. Georgeot, O. Giraud, Distinguishing humans from computers in the game of go: a complex network approach. Europhys. Lett. **119**(4), 48001 (2017)
10. D. Benson, *Music: A Mathematical Offering* (Cambridge University Press, Cambridge, 2006)
11. W.B. De Haas, J.P. Magalhães, F. Wiering, R.C. Veltkamp, Automatic functional harmonic analysis. Comput. Music. J. **37**(4), 37–53 (2013)
12. P. Kröger, A. Passos, Sampaio, M. De Cidra, G.: In: *ICMC* (2008)
13. M. Rohrmeier, I. Cross, In: *10th International Conference on Music Perception and Cognition*, vol. 6. (Hokkaido University Sapporo, Japan, 2008), pp. 619–627
14. M. Neuwirth, D. Harasim, F.C. Moss, M. Rohrmeier, The annotated beethoven corpus (abc): a dataset of harmonic analyses of all beethoven string quartets. Front. Digital Human. **5**, 16 (2018)
15. J. Hentschel, M. Neuwirth, M. Rohrmeier, The annotated mozart sonatas: score, harmony, and cadence. Trans. Int. Soc. Music Inf. Retrieval **4**(1), 67–80 (2021)
16. A. Marsden, Schenkerian analysis by computer: a proof of concept. J. New Music Res. **39**(3), 269–289 (2010). https://doi.org/10.1080/09298215.2010.503898
17. A. Elgammal, Music scholars and computer scientists completed Beethoven's Tenth Symphony aided by machine learning. https://tinyurl.com/f7e2mkea
18. K. Landsnes, L. Mehrabyan, V. Wiklund, F.C. Moss, R. Lieck, M. Rohrmeier, In: *16th Sound & Music Computing Conference* (2019), pp. 250–254
19. D.W. Beach, The origins of harmonic analysis. J. Music Theory **18**(2), 274–306 (1974)
20. O. Darrigol, The acoustic origins of harmonic analysis. Arch. Hist. Exact Sci. **61**(4), 343–424 (2007)
21. F.C. Moss, M. Neuwirth, D. Harasim, M. Rohrmeier, Statistical characteristics of tonal harmony: a corpus study of Beethoven's string quartets. PLoS ONE **14**(6), 0217242 (2019)
22. L.G. Ratner, Key definition, a structural issue in Beethoven's music. J. Am. Music. Soc. **23**(3), 472–483 (1970)
23. K.M. Knittel, Wagner, deafness, and the reception of Beethoven's late style. J. Am. Music. Soc. **51**(1), 49–82 (1998)
24. M.E. Bonds, Irony and incomprehensibility: Beethoven's "serioso" string quartet in f minor, op. 95, and the path to the late style. J. Am. Music. Soc. **70**(2), 285–356 (2017)
25. D. Tymoczko, The geometry of musical chords. Science **313**(5783), 72–74 (2006)
26. M. Buongiorno Nardelli, *The hitchhiker's guide to the all-interval 12-tone rows*. arXiv:2006.05007 (2020)
27. M. Buongiorno Nardelli, Tonal harmony and the topology of dynamical score networks. J. Math. Music 1–15 (2021)
28. M. Neuwirth, D. Harasim, F.C. Moss, M. Rohrmeier, https://github.com/DCMLab/ABC

29. G. Zipf, *The Psycho-biology of Language: An Introduction to Dynamic Philology*, 1st edn. (The MIT Press, Cambridge, USA, 1935)
30. MuseScore. version 3.6.2, released under the GNU GPLv2 license
31. S. Brin, L. Page, The anatomy of a large-scale hypertextual web search engine. Comput. Netw. ISDN Syst. **30**(1–7), 107–117 (1998)
32. L. Ermann, K.M. Frahm, D.L. Shepelyansky, Google matrix analysis of directed networks. Rev. Mod. Phys. **87**(4), 1261 (2015)
33. M.E. Newman, M. Girvan, Finding and evaluating community structure in networks. Phys. Rev. E **69**(2), 026113 (2004)
34. S. Fortunato, Community detection in graphs. Phys. Rep. **486**(3–5), 75–174 (2010)
35. A.-L. Barabási, R. Albert, Emergence of scaling in random networks. Science **286**(5439), 509–512 (1999)
36. R. Albert, A.-L. Barabási, Statistical mechanics of complex networks. Rev. Mod. Phys. **74**(1), 47 (2002)
37. H. Ebel, L.-I. Mielsch, S. Bornholdt, Scale-free topology of e-mail networks. Phys. Rev. E **66**(3), 035103 (2002)
38. D. Donato, L. Laura, S. Leonardi, S. Millozzi, Large scale properties of the webgraph. Eur. Phys. J. B **38**(2), 239–243 (2004)
39. G. Pandurangan, P. Raghavan, E. Upfal, Using pagerank to characterize web structure. Internet Math. **3**(1), 1–20 (2006)
40. O. Giraud, B. Georgeot, D.L. Shepelyansky, Delocalization transition for the google matrix. Phys. Rev. E **80**(2), 026107 (2009)
41. B. Georgeot, O. Giraud, D.L. Shepelyansky, Spectral properties of the google matrix of the world wide web and other directed networks. Phys. Rev. E **81**(5), 056109 (2010)
42. V.D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp. **2008**(10), 10008 (2008)
43. T. Hedges, M. Rohrmeier, In: *International Conference on Mathematics and Computation in Music*. (Springer, 2011), pp. 334–337
44. C. Weiß, F. Zalkow, V. Arifi-Müller, M. Müller, H.V. Koops, A. Volk, H.G. Grohganz, Schubert winterreise dataset: a multimodal scenario for music analysis. J. Comput. Cult. Heritage **14**(2), 1–18 (2021)
45. E. Anzuoni, S. Ayhan, F. Dutto, A. McLeod, F.C. Moss, M. Rohrmeier, In: *18th Sound and Music Computing Conference* (2021), pp. 284–291
46. M. Rohrmeier, In: *21st Int. Society for Music Information Retrieval Conference* (2020)
47. T. Pankhurst, *SchenkerGUIDE: A Brief Handbook and Website for Schenkerian Analysis* (Routledge, New York, 2008)