Visualizing Music Information: Classical Composers Networks and Similarities

Patrick Georges[1]
[1] Graduate School of Public and International Affairs, University of Ottawa, Social Sciences Building, Room 6011, 120 University Ottawa, Ontario K1N 6N5 Canada

Abstract
This paper illustrates different information visualization techniques (data visualization) applied to a classical composers' database. In particular we present composers network graphs, heat maps and multidimensional scaling maps (the latter two obtained from a composer distance matrix), composers' classification maps using support-vector machine and K-Nearest Neighbors algorithms, and dendrograms. All visualization techniques have been developed using Python programming and libraries. The ultimate objective is to enhance basic music education and interest in classical music by presenting information quickly and clearly, taking advantage of the human visual system's ability to see patterns and trends.

Key Words: Network Graphs, Heat-Maps, Multidimensional Scaling, Support-Vector Machines, K-Nearest Neighbors, Dendrograms, Music Information Retrieval, Python programming.

1. Introduction

This paper illustrates different information visualization techniques (data visualization) applied to a classical composers' database, *The Classical Music Navigator*, hereafter referred to as *CMN*, a website created by Charles H. Smith (2000), and available at http://people.wku.edu/charles.smith/music/. In particular we present composers network graphs, heat maps and multidimensional scaling maps (the latter two obtained from a composer distance matrix), composers' classification maps using support-vector machine and K-Nearest Neighbors algorithms, and dendrograms. All visualization techniques have been developed using Python programming and libraries. The ultimate objective is to enhance basic music education and interest in classical music by presenting information quickly and clearly, taking advantage of the human visual system's ability to see patterns and trends.

Khulusi et al. (2020) have recently surveyed a large amount of the literature that focuses on the unique link between musicology and visualization by classifying 129 related works according to the visualized data types and analysing which visualization techniques were applied for certain research inquiries. The survey covers visualization of musical scores, visualization of musical sound, visualization of musical collections (including classification, recognition, annotation, and the retrieval of music), visualization of musicians (including composers and singers, but also instrument makers, etc.), their biographical information and similarities, and visualization of instruments (including how instruments operate). The intersection of musicology and visualization brings a diversity of innovative applications designed for a variety of purposes. As Khulusi et al. (2020) explain, on the one hand, musicologists are served with interactive tools to analyze musicological data, and on the other hand, applications are tailored for the broad public to communicate and to teach aspects of music in a more intuitive, playful manner.

Smith (2000) created the *CMN* as a reference work and experiment in music education. To serve this music education objective, the *CMN* consists of five compilations of material: (1) an alphabetically-arranged, 'Composers' list containing basic data, major works, and influences of 500 individuals; (2) a 'Basic Library' list of works culled from this composer list (and re-arranged by musical genre); (3) a 'Geographical Roster' in which the names of the 500 composers are listed under the names of the countries with which they were (/are) associated; (4) an alphabetically-arranged 'Index of Forms and Styles' listing the names of composers associated with each subject entry; and (5) a 'Glossary,' which defines terms used in the *CMN*. One objective of the *CMN* website was, from the very onset linked, at least implicitly, to early efforts in music information retrieval (MIR). For example, the *CMN* site explains that many introductions to the classical music world are in the business of inculcation through lists of 'mandatory' composers and compositions to explore. Yet, most people explore new subjects by starting with the familiar, and in the case of music, this may mean hearing a composer or a composition that one likes and then searching for more music of the same type. The site gives the following example:

> "Suppose you hear the Ravel G major piano concerto on the radio, and take an immediate liking to it. Our database will help you extend this interest to other music by making it possible for you to quickly identify: additional works by Ravel, other piano concerti, other works for piano in general, other concerti in general, composers allied to the same general period and style (Impressionism) as Ravel, other French composers, composers and styles that influenced Ravel, and composers influenced by Ravel."

Thus, the *CMN* anticipated the general idea of a recommender system that is now commonly and automatically implemented in Pandora, Spotify, Last.fm, YouTube and other music streaming platforms that have algorithms proposing what an auditor may want to listen next. These algorithms and their improvement are largely tributary to the field of MIR, which develops innovative content-, context- and user-based searching schemes, music recommendation systems, and novel interfaces to make the vast store of music available to all.[1]

Perhaps one obstacle in the *CMN* original objective of music education is, at the onset, a lack of any supporting tools enhancing the human visual system's ability to see pattern and trends. This paper explores some visual tools that can support the education mission of a music database such as the *CMN* (or any other music database for that matter). First, given the network of influences of composers assembled in the *CMN,* methods developed in network visualization (or standard graph theory) seem appropriate. Hence, Section 2 of the paper illustrates how to apply network visualization techniques to the *CMN* database.[2] In Section 3 of the paper, we take on the challenge of detecting similarities across composers, an explicit objective of the *CMN*, and explain the methodology underlying the construction of composers' similarity indices (based on a cosine similarity measure). Even with a relatively small database of just 500 composers, this leads to 250,000 (500x500) bilateral

[1] One of the earlier survey articles on MIR is Orio (2006). Schedl et al. (2014) provide a survey of more recent developments and applications.
[2] For papers that also present social networks in different 'music worlds', see Crossley et al. (2015). For a social network analysis of British composers from 1870, see McAndrew and Everett (2015).

indices of similarity.[3] The sheer dimension of this information prevents easy reporting in an academic paper. But Section 3 shows how this information can be captured visually in one graph, using a heat map. Furthermore, this index of similarity is now accessible in the *CMN* website in the form of lists of 15 most similar composers to each subject composer. With open access to the composers' similarity index, researchers may use it either in their own research, or as a benchmark for purposes of comparison to, say, similarity indices extracted from audio files, or with alternative data used in MIR research. In passing, this shows that context-based MIR (i.e., MIR based not on audio files but on more general information on composers including their 'cultural' context – and which underlies the *CMN* methodology and philosophy) – remains useful to capturing similarities across composers.[4] Section 4 pursues with Multidimensional Scaling (MDS), a technique that transforms the composers' similarity/distance matrix (from Section 3) into visual (MDS) maps. It applies support-vector machine and K-Nearest Neighbors algorithms to classify composers into several classes. It also uses hierarchical clustering analysis to produce dendrograms of composers. The final section concludes and discusses issues related to music discovery, serendipity, semantic labeling of music and artificial intelligence in music.

2. Composers' visualization networks

We begin with some definitions and terminology on networks taken from standard graph theory. We use the notations and some definitions from Jackson (2011). A network is represented as a graph on a set N of nodes (sometimes referred to as vertices), with a finite number of members n. A graph or network is a pair (N, g), where g is an $n \times n$ adjacency matrix on the set of nodes, where $g_{i,j}$ indicates the relationship between nodes i and j. Here I focus on cases where $g_{i,j} \in \{0,1\}$ so that a relationship is either present ($g_{i,j} = 1$) or absent ($g_{i,j} = 0$). A graph is undirected if g is required to be symmetric so that $g_{i,j} = g_{j,i}$, and is directed otherwise. Whether or not a network is directed or undirected depends on the application. In the composer database where an influence from a composer i to a composer j exists then $g_{i,j} = 1$, but if the influence is not reciprocal (j did not influence i) then $g_{j,i} = 0$. The relationship between two nodes i and j, where $g_{i,j} = 1$ is referred to as an edge (or sometimes link or tie) and in our case of directed network, a directed edge.

A (directed) walk in a network (N, g) refers to a sequence of nodes, $i_1, i_2, i_3, \ldots, i_{K-1}, i_K$ such that $g_{i_k,i_{k+1}} = 1$ for each k from 1 to K. The length of the walk is the number of links in it, or $K - 1$. A (directed) path in a network (N, g) is a walk in (N, g), $i_1, i_2, i_3, \ldots, i_{K-1}, i_K$, such that all the nodes are distinct. The geodesic distance between two nodes of a (directed) network is the length of a shortest (directed) path between them.

The neighbors of a node i in a undirected network (N, g) are denoted $N_i(g) = \{j, g(i,j) = g(j,i) = 1\}$. The degree of a node i in the undirected network (N, g) is the number of neighbors that i has in the network, so that $d_i(g) = N_i(g)$. For directed networks we also introduce 'in-degree' and 'out-degree' definitions. The out-degree of a node i is the number $d_i^{out}(g) = \{j, g(i,j) = 1\}$, that is, the number of reported influences of i on all j. The 'in-degree' of a node i is the number $d_i^{in}(g) = \{j, g(j,i) = 1\}$, that is, the

[3] If we abstract away from computing the similarity of a composer with him/her-self, then we strictly have n*(n-1) = 500x499 = 249500 indices.
[4] For a recent text on audio/content-based MIR, see Müller (2015). For a text that puts more emphasis on alternative 'non-audio'/context-based MIR, see Knees and Schedl (2016).

number of reported influences of all *j* on *i*. We may possibly define the degree of *i* in the directed network as given by: $d_i(g) = d_i^{out}(g) + d_i^{in}(g)$.

There are several algorithms that can be used to visualize a network through a pictorial representation of the nodes and edges of the network (*N, g*). There can be very different layouts or representations of the network itself depending on the algorithms used. As an illustration in Section 2, we use the ForceAtlas2 algorithm (Jacomy et al. (2014), a forced-based layout whereby the algorithm modifies an initial (random) node placement by continuously moving the nodes according to a system of forces based on a metaphor of springs and electric charged particles. The 'spring-electric' layout uses the attraction formula of springs (between nodes connected with an edge) and the repulsion formula of electrically charged particles (between any nodes). It uses the attraction force (or restoring force) formula of springs ($F_a(i_1,i_2) = k_a dist(i_1,i_2)$): the more you stretch something, the harder it becomes to keep stretching. Or as you stretch something out, there is a restoring force (of opposite sign) that you have to compete with.[5] Thus, connected nodes with closer geometric distance *dist*() attract less (the restoring force is lower) than for more distant connected nodes. It also uses the repulsion formula of electrically charged particles (electrons), $F_r(i_1,i_2) = k_r / (dist(i_1,i_2))^2$ where *dist*() is the geometric distance between two nodes (charges) so that closer entities repulse more. Hence the spring-electric analogy suggests that closer entities/nodes attract less but repulse more. These forces create a movement that converges to a balanced state.

The main difference between several force-based algorithms is in the actual value of the exponent associated with *dist*() in the attraction and repulsion (*a*, *r*) formulas (with a '-' sign if *dist* is in the denominator). For example, the spring-layout analogy explained above is (*a, r*) = (1, -2), the Fruchterman and Rheingold algorithm is (2,-1) and the ForceAtlas2 algorithm is (1, -1) where in this case the attraction force is the one given above, [6] while the degree-dependent repulsion force is given by: $F_r(i_1,i_2) = k_r \dfrac{(d_{i_1}+1)(d_{i_2}+1)}{dist(i_1,i_2)}$ where d_i is the degree of a node *i* as explained above. Hence the repulsion force is proportional to the degrees (plus 1) of the two nodes. The ForceAtlas2 algorithm tends to increase the repulsion between highly connected nodes and to produce a lesser repulsion force between poorly connected nodes and highly connected ones, and an even lesser repulsion force between poorly connected nodes. This avoids the cluttering effect of some algorithms in which a forest of leaves (poorly connected nodes) surrounds the few highly connected nodes.[7]

[5] The spring constant, k_a, is associated with a negative sign if we are talking of the restoring force, because the restoring force is in the opposite direction to the extension. If F_a is the force we apply, then the negative sign goes away.

[6] There is an alternative option (dissuade hubs) in ForceAtlas2 given by $F_a(i_1,i_2) = k_a dist(i_1,i_2) / (d_{i_1}^{out}+1)$ whereby, if node i_1 is an 'hub' (i.e., a node with a high out-degree, i.e., many arrows or influences pointing to other nodes) then this option will tend to reduce the hub's attraction toward other connected nodes (i_2), pushing hubs towards the periphery while keeping 'authorities' (nodes with a high in-degree but low out-degree) in the center.

[7] The user can choose the value of k_r which provides a scaling effect. The higher k_r the larger the graph will be.

Figure 1 provides an illustration of the composers' network.[8] A color code has been introduced so that nodes in blue represent composers from the Medieval and Renaissance periods; green represents Baroque; red is Classical; cyan is Romantic; and magenta represents composers of the 20th century. Of course, instead of the period, we can pick colors to illustrate other features of the composer such as, say, the country of origin. By displaying the number of influences of a composer (known as the out-degree as explained above) as the size of the nodes, the visualization also shows that there are a few nodes with a lot of influences. These large nodes are known as 'hubs' – composers who have influenced many other composers in the network. To avoid cluttering, we only labelled the top-100 ranked composers. What is remarkable in Figure 1 is the accuracy of the composers' localisation on the map where composers from the same period appear closer together reflecting a stronger intra period network of influences. To underline this fact, it is perhaps relevant to show in Figure 2 the little visual information we get with the Fruchterman-Rheingold algorithm which clutters all hubs in the centre of the graph and offers no hint of the existence of clusters by period.[9]

It is sometimes said that these social network graphs all look similar. However, Figures 1 and 2 illustrate the importance of choosing the appropriate algorithm so as to provide useful visualization information to the user. This said, quantitative metrics remain essential to shed further information that would be quite difficult to extract from the initial *CMN* database of direct influences or even from the graph in Figure 1.[10] The density of network is the ratio of actual edges in the network to all possible edges (an arrow between any two nodes). Hence network density (a number between 0 and 1) gives a quick sense of how closely knit the network is. In our case, we have 500 nodes and 3724 edges, so that the composer network density is 0.0149. Hence, this classical composer network is on the lower edge of the (0, 1) range, but still far from 0.

Another structural metric of a network is the concept of triadic closure. It supposes that if two people know the same person, they are likely to know each other. Adapted to the composer network, if composers j and k have been influenced by i, then perhaps k was also influenced by j (or j by k). Or if i and j have influenced k, then perhaps i has also influenced j (or j has influenced i). One way of measuring triadic closure is through the concept of transitivity, the ratio of all existing triangles (between three composers) over all potential triangles. Thus, like density, transitivity expresses how interconnected a network is, and is represented by a number between 0 and 1. In our case, we obtain a value of 0.0537. Because the graph is not very dense, there are fewer potential triangles, which may result in a slightly higher number than the one obtained for density (0.0149). A third structural measure of interest is the shortest path measurement which calculates the shortest possible series of nodes and edges that stand between any two nodes. In absence of direct influence of composer i on k, but if i has influenced j and j has influenced k, then the shortest path length between i and k is 2. A short-path of length 2 may lead to conjecture about whether

[8] Figures 1 to 6 have been generated using the Python library NetworkX and dependencies. With the exception of Figure 2, all figures also use the ForceAtlas2 algorithm implemented for NetworkX by Bhargav Chippada (2017).
[9] This results from the fact that *a-r* (= 2) in ForceAtlas2 is less than *a-r* (= 3) in Fruchterman and Rheingold. Visual clusters denote structural densities (the ratio of actual edges on potential edges) when *a-r* is low, that is when the attraction force depends less on distance and when the repulsion force depends more on it. We introduce more formally the notion of a network density in the next paragraph.
[10] Ladd et al. (2017) discuss some of these metrics and explore in details how to compute them using Python NetworkX library.

or not an 'indirect' or residual musical influence of *i* exists on the music of *k,* which might be of interest to a musicologist. We can obtain the shortest path metric for any two composers *i* and *j* in the database. For example, the shortest path between Lasso (Renaissance) and Debussy (transition from late romantic to early modern) is represented by the series of nodes: [' Lasso', ' Charpentier, M-A', ' Couperin, F', ' Debussy'] (path length = 3). The shortest path between Monteverdi (end of Renaissance, early Baroque) and Glass (20th century) is: [' Monteverdi', ' Schutz', ' Bach, JS', ' Debussy', ' Glass'] (path length = 4). The shortest path between Byrd (Renaissance) and Gorecki (20th century) is: [' Byrd', ' Purcell', ' Handel', ' Beethoven', ' Gorecki'] (path length = 4). Interestingly, although we might expect shorter shortest paths among contemporaneous composers, this is not necessarily the case. For example, the shortest path between Elgar (born 1857) and Debussy (born 1862) is [' Elgar', ' Walton', ' Britten', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy'] and the shortest path between Vaughan Williams (born 1872) and Debussy is [' Vaughan Williams', ' Holst', ' Britten', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy'], in both cases a path length of six steps. Of course, in a directed network, we must also expect that no shortest path exist between composers born much later than Debussy and Debussy (even if Debussy has influenced the much younger composer). In this case, we could more appropriately compute all direct and indirect influences (all shortest paths) of Debussy on any composer *j*. In appendix 1, we list all shortest paths between any composers *i* and Debussy. A path length of 1 is a direct influence on Debussy as recorded in the *CMN*. Note that the longest shortest paths are 7 steps and there are four of them. Six degrees of separation, the idea that all people are six or fewer connections away from each other is therefore nearly fulfilled for Debussy.[11]

Beyond some structural measures of the composers' network, we can also discuss which nodes are the most important in the network through measures of centrality, in particular, degree, betweenness centrality and eigenvector centrality. 'Degree' is the most common way of finding important nodes. Table 1 sorts composers by their out-degree. Recall that the out-degree of a node represents the number of composers the node (the subject composer) has influenced. It is thus a measure of how influential a composer has been in the history of classical music. Table 1 reports only the 20 most influential composers. Results in Table 1 are not surprising. We have one Baroque composer (JS Bach), three classical (Beethoven, WA Mozart, J Haydn), 10 Romantic (of which Wagner, Brahms, Liszt) and six modern (of which Stravinsky, Schoenberg and Bartok), all of them known to have had a large and sustained influence on many composers during their life and beyond. Table 2 provides the in-degree of a composer, that is, the number of composers who influenced him/her, transforming the latter into a 'sink' of influences. Of course, there are different types of sinks. A composer *i* who has been influenced by many other composers may be seen as prestigious or knowledgeable ('Authorities') and some (younger?) composers may seek to be influenced by *i* for that very reason. But sinks may also reflect information overload or noise and interference due to the contradictory messages from different sources and in this case (younger?) composers might not want to seek influence from them. Composers with a high in-degree who also have a high out-degree may be viewed as 'communicators' or 'facilitators' of the network (e.g., Debussy, JS Bach, WA Mozart, Ravel, Liszt, Wagner, and Stravinsky). They are influencers building on the shoulders of others. Composers with a high in-degree but a low out-degree are pure information sinks (they do not share or transfer the knowledge they may have learned from others). It may be because of 'demand', 'supply' or 'time constraint' effects. Some

[11] Note however that out of the 499 composers (besides Debussy) in the database, there are 243 composers for which there is no shortest path to Debussy.

composers, as said above, are a mediocre source of influence because they are viewed as too indiscriminate in their own sources of influence, resulting in a low demand for their influence. The supply effect may reflect voluntary isolationism and lack of interest in transmitting knowledge. Finally, according to the time constraint effect, composers from a more recent period may not have had the time needed to become a source of influence. Observe for example in Table 2 that 11 of those 20 sinks of influences are from the 20th century period (e.g., Ligeti, Ginastera, Crumb, Penderecki) and among them Ligeti is an important sink of influences but has not yet influenced others (in-degree = 19, out-degree =1) because of the time constraint effect (or database limitation, e.g, no records on younger active composers). Perhaps Britten is an example of a sink seeking some isolationism and who did not bother to transmit knowledge (in-degree = 20, out-degree = 8) while Sullivan is another British example of a sink from the Romantic period (in-degree = 16, out-degree = 2) that had little influences and might have suffered from his reputation among the musical establishment of writing frivolous music. Composers with low in-degree but high out-degree are perhaps best described as outsiders and innovators, pushing for new developments without necessarily relying on many previous influences. Composers such as Haydn, Beethoven, Berlioz, Schumann and Schoenberg fit this description to some extent. Finally, some composers have low in-degree and low out-degree, composing out-of-the-loop, at the periphery. Karl Stamitz (the elder son of Johann Stamitz) fits the description. He composed some orchestral works and chamber music that stylistically resembles that of Mozart and Haydn, visited many cities as a virtuoso on the violin and viola, but never managed to gain a permanent position with a European court or in one of the orchestras of his time. He finally moved to Jena, a city in central Germany, where there was neither a town band nor an orchestra. The table below summarizes this discussion.

Summary table

	High out-degree (d_i^{out})	**Low out-degree (d_i^{out})**
High in-degree (d_i^{in})	**'Authorities' and highly influential composers: 'Communicators'** e.g.: Debussy, JS Bach, WA Mozart, Ravel, Liszt, Wagner, Stravinsky	**Information Sinks** e.g.: Britten (in-degree = 20, out-degree = 8) (supply effect) Sullivan (in-degree=16, out-degree =2) (demand effect) Ligeti (in-degree = 19, out-degree =1) (Time-constraint effect)
Low in-degree (d_i^{in})	**Outsiders and Innovators** e.g.: Haydn, Beethoven, Berlioz, Schoenberg	**Out-of-the-loop, at the periphery** e.g.: Karl Stamitz

Two other measures of centrality (importance of a node) are betweenness centrality and eigenvector centrality. Betweenness centrality tries to capture nodes that are important not because they have many out-degrees or in-degrees, but because they stand between groups, giving the network connectivity and cohesion. If a composer happens to often be on shortest paths (as defined above) between any two composers, they will score high on betweenness centrality. (In the list of shortest paths between any *i* and Debussy in Appendix 1, we can see some composers who appear more frequently.) Observe in Table 3 that composers with high betweenness-centrality are present in all major periods:

Renaissance: Sweelinck and Palestrina; Baroque: Schütz, Purcell, JS Bach and Handel; Classical: Mozart and Beethoven; Romantic: Chopin, Wagner, Liszt, Debussy, and Modern: Ravel, Schoenberg, Bartok, Stravinsky, Britten, Gershwin and Shostakovich. Finally, eigenvector centrality cares not only about the number of connections (influences) but also the 'quality' of these connections (i.e., whether your connections are also well connected). Table 4 provide a list of such composers. All of them are from the Modern period, which reflects the advantage of coming chronologically later. Modern composers have the opportunity to cherry-pick among an increasingly large pool of influencers from several periods, an opportunity that a Renaissance or Baroque composer did not have. In Table 4, we see that Ligeti (1923-2006) has been influenced by a range of important and influential composers from late Renaissance to the modern period (Ockeghem, Monteverdi, JS Bach, R. Schumann, Liszt, Debussy, Stravinsky, Bartok, Kodaly, Berg, Webern, Boulez, Cage, Reich, Riley, Stockhausen, Nancarrow, and Varèse).

A characteristic of the composers' network studied above is that it connects composers of very different periods, ranging from Medieval to Modern composers. Yet 'social' networks are often used to study connections between people alive during one single period. Half the *CMN* database is made up of composers born in 1862 and after (249 composers out of 500). Hence, we still have a decent size 'social' network when concentrating on this period alone. The choice of 1862 is somewhat arbitrary but justified here by the fact that Debussy was born in 1862 and, as discussed in Griffiths (1978) in his '*Concise History of Modern Music: From Debussy to Boulez*', although he might be classified as 'late Romantic', he was instrumental in transforming classical music towards the modern idiom of the 20th century through his impressionist style.

The 20th century music network is somewhat biased towards American and British composers (with respectively 77 and 29 composers), while there are 31 French composers, 23 Germans/Austrians, 16 Italians, 16 Russians, 15 'Central'-Europeans (of which 6 Hungarians, 5 Bohemians/Czech and 5 Polish), 12 North-Europeans composers (3 Danish, 3 Swedish, 4 Finnish, 2 Estonians), 12 South-Americans (4 Mexicans, 3 Argentinians, 2 Brazilians, 1 Paraguayan, 1 Venezuelan, and 1 Cuban), 8 Spanish, and 10 composers of other nationalities (2 Dutch, 2 Swiss, and one each from Belgium, Greece, Canada, Australia, Japan, and China). Of course, some composers have had several citizenships over their lifetime and in this case, we selected just one citizenship depending on the biography and most relevant musical context of the composer (so as to be able to attribute a single colour by node in the network as explained shortly).[12] The adjacency matrix is now of dimension 249x249. Some 'modern' composers have been influenced by composers from earlier periods, but we disregard this fact to only take into account the network of composers within the 'modern' period. The number of nodes is 249 and the number of edges is 1325, hence the average in/out degree of a node is 5.3. Density of the network is 0.02146 (higher than the full composers' network) and the triadic closure is also higher at 0.03596. Hence this '20th century/modern' network is denser than the full

[12] For examples, Varèse (1883-1965) was a French-born composer who spent the greater part of his career in the United States from 1915 onwards and took the American citizenship in 1927. He is thus grouped here with American composers (in red). Xenakis (1922-2001) was a Greek composer who moved to France in 1947 and become a naturalised citizen in 1965. On balance, he was generally more entwined with French than Greek cultural life and thus we group him with French composers (in blue). Nancarrow (1912-1997) was born in the USA and became a Mexican citizenship in 1956 but he leaved in relative isolation in Mexico while his musical influences are American. Thus we group him with Americans (in red).

network analyzed above. This reflects that there is, on average, more documented influences per composers of the 20th century than for the entire network.

We have experimented with node features by color coding nodes according to citizenship (or regions of affinity) in Figure 3(a), by composer's main style (explained shortly) in Figure 3(b). Figure 3(c) zooms Figure 3(b) in the middle of the graph to reduce the effect of names' overlapping. We also produced a setting with gender and race colour-coding in Figure 3(d).[13] We use a pseudorandom number generator to obtain an initial placement of nodes (i.e., a placement that is statistically random but created in a deterministic manner). This ensures the same final 'equilibrium' placement of nodes across maps. Unlike composers from earlier periods, quite a few twentieth century composers are little-known to the public at large and Figures 3(a-d) may help discover (and hopefully listen to) some composers by exploring connections or influences pointing to them from the largest nodes (typically the most well-known composers in this period) whose size depends on the out-degree (influence) of a composer.

First, notice that Figures 3(a-b) illustrate the distinct paths adopted by 20th century composers as if there were 'two centuries in one'. As Pauls (2014) puts it: "An outstanding feature of the twentieth-century has been the divergence of European 'art' music into two general areas which do not overlap to the same extent that they do in previous centuries. That is, the performing repertoire is at odds, sometimes dramatically so, with a competing canon of works considered to be of greater importance from an evolutionary historical point of view". The 'two centuries in one' feature can roughly be seen in Figure 3(a-b) where composers in the bottom (South) part of the map make up the bulk of the performing repertoire of this century classical music, pursuing (to some extent) the romantic style of the 19th century, and, more generally, pursuing the five pillars of the 'Common Practice Period' of Western classical music (1600-1900), *Tonality*, *Vocabulary*, *Texture*, *Sonority*, and *Time*.[14] As we move towards the most Northern part of the map, however, we find composers who have completely changed the musical elements of the 'Common Practice Period' and are often viewed (loosely) as the 'avant-garde' of the music of their time. Magnuson (2008) offers an interesting discussion about which of the five pillars of the

[13] We have also produced a setting with age groups. See Graphical Appendix, Figure A1, where age groups are color-coded as cyan for the late Romantic composers; magenta for those born 1870-1899; blue for those born 1900-1929; and red for those born in 1930 and after. Those age groups are relatively arbitrary but a 30-year length may be thought of as 'generations', and those groups as overlapping generations of composers active in late 19th century and in the 20th century. This setting seems to suggest that older age groups are localised in the West, South and East parts while younger generations (those who changed most pillars of the Common Practice Period) are closer to the center and towards the North/Northeast. This also matches the description in Figure 3(b).

[14] Magnuson (2008) describes the five pillars of the Common Practice Period as follows: 1. *Tonality*—The essential organisation around a single pitch, the tonic, which provides a home base to the ear; 2. *Vocabulary*—A diatonic pattern of seven stepwise pitches called major and minor scales; 3. *Texture*—Texture of the Common Practice Period is created with counterpoint, which is two or more simultaneous individual and independent lines, each of which confirms the pre-eminence of tonic and utilizes the vocabulary of a major or minor scales. 4. *Sonority*—consonant sonority of the Common Practice Period is a group of three notes arranged in thirds (tertian triad). Dissonance can be used on occasion in the form of a group of four notes arranged in thirds (tertian tetrad); and 5. *Time*—The essential time organization of the Common Practice period is based on a consistent and unchanging beat. These beats organize into 2, 3, or 4 essential pulses per measure, with the first beat always the strongest. Each beat can sub-divide into two parts (simple meters) or three parts (compound meters).

Common Practice Period have been basically maintained, generally modified, or completely modified in the different styles of 20th century classical music (e.g., impressionism, primitivism, neoclassicism, expressionism, serialism, indeterminism, minimalism, neo-romanticism, etc.). He assumes that when a composer either generally modifies or completely changes more than one of these five elements, then, a new music (or Uncommon Practice) is created. Twentieth century styles are discussed shortly later.

Second, observe in Figure 3(a) that the placement of nodes remains dependent on 'citizenship' despite so-called 20th century globalisation. Without going into much detail, we see Russian composers (in cyan) in the South-East of the map (e.g., Rachmaninov, Medtner, Miaskovsky, Glière, Glasunov, Khachaturian, A. Tcherepnin, Shchedrin, Prokofiev, and slightly further to the North, Stravinsky and Shostakovich). In the East we encounter North-Europeans (Swedish, Norwegian, Danish, to which we add Finnish and Estonian) in deepskyblue (e.g., Stenhammar, Sibelius, Nielsen, Holmboe, Norgard, Sallinen, Larsson, Tubin, and in direction of the center, Kokkonen, Part, Rautavaara, etc.). We see British composers (in hotpink) positioned in the South/South-East (Butterworth, Vaughan-Williams, Holst, and, starting from Holst and moving in the Northward direction, Howells, Finzi, Moeran, Rutter, Bax, Alwyn, Walton, Berkeley, Tippett, Rubbra, and further to the center, Britten and Warlock. Also starting from Holst we see a strand of British composers in the Westward direction: Delius, Clarke, Ireland, Bridge, Bliss). In the South-West we see a series of Spanish composers (in yellowgreen) (Granados, Falla, Monpou, Turina, Rodrigo) and Italian composers (in magenta), Giordano, Mascagni, Respighi, Castelnuovo-Tedesco, and further West, Wolf-Ferrari, Cilea, Pizzetti. In the West/South-West, we see French composers (Vierne, Langlais, Dupré, Tournemire, Duruflé, Pierné, Lily Boulanger, and further East, Debussy, Dukas, etc.) In the East we encounter German composers (in yellowgreen) (e.g., Pfitzner, R. Strauss, and further East Reger, Karg-Elert and Orff, and Austrian composers in darkgreen (e.g., Zemlinsky, Schreker, Schoenberg, Berg, Webern). American composers firmly occupy the center and the Northeast side of the map. Note that this clustering of composers by citizenship might, but does not necessarily, imply a geographical clustering (after all composers have migrated without necessarily adopting a new citizenship!). It may however be interesting to see whether the historical unfolding of the twentieth century music is concomitant to geographical shifts in cultural centers.

An overemphasis on nationality cannot capture the rich network of the 20th century composers (and their music). Figure 3(b) present the main styles of music of 20th century composers. Associating a unique or main style/color to one composer is often a very restrictive and misleading assumption as many composers explored different styles over their lifetime. The *CMN* provides a list of styles for most composers. When only one style was provided, this style was attributed to the composer. When more than one style was provided, an additional research was done by reviewing the bios of composers in several sources such as Wikipedia and other composers websites to ascertain which style seemed to dominate (in terms of number of years or overall influence) during the composer lifetime. We however, fully appreciate the limit of this approach and Section 3 will deal more appropriately with this issue. But for the time being we deliberately take this short-cut to explore main styles over the 20th century. According to Magnuson (2008), music of the 20th Century is unique in its pluralism. "Composers began to explore a more personal and individual approach to music creation, forming their own microcosms or 'small universes'. No longer bound to the rules formed by one musical approach, they customized sound to suit their own views and preferences." For Magnuson, there were three important small universes or microcosms near the turn of the 20th century: Impressionism (Debussy, Ravel in the South-West in Figure 3(b)), Primitivism (Stravinsky, Bartok in the East and Center), and Expressionism (Schoenberg, Berg, Webern—in the Northwest). Impressionism (as represented by Debussy and Ravel and other composers in blue in the South/Southwest)[15]was a reaction to the state of music at the end of the 19th century, that is, late Romantic composers who lived well into the 20th century and who essentially belong to the Common Practice Period even if they made some concessions to the new century. These essentially Romantic composers are located in the East/Southeast and West/Southwest parts of Figure 3(b) and are represented in deepskyblue: (e.g., R. Strauss, Reger, Rachmaninov, Glazunov, Vaughan Williams, Sibelius, Nielsen, etc.).

Expressionism followed the path of the common practice period but completely mutated its basic pillars of tonality, vocabulary, texture, sonority and time. Besides Schoenberg, Berg and Webern, other composers pursued expressionism. These composers gravitate not far away from Schoenberg, Berg and Webern in Figure 3(b), e.g., Krenek, Wolpe, Henze, Husa, Kraft, Gerhard, Toch, KA Hartmann, Carter. Expressionism itself led to Serialism (some representative composers, at least during a part of their life, are given in pink, e.g. Dallapiccola, Davidovsky, Riegger, Sessions, Eisler, Skalkottas, Petrassi, Finney, Babbitt, Rochberg, Walker, Perle, Schuller, Musgrave, Druckman, Birtwistle, Davies, R.R. Bennett, Wuorinen, Tower, Zwilich).

Primitivism positioned itself somewhere between Impressionism and Expressionism, and eventually led to 1) Neo-Classicism (composers in dark green in Figure 3(b)[16], essentially grouped in the East part of the graph, and to 2) the revival of Nationalism as a source of inspiration, a trend that began with Glinka and Dvorak in the 19th century (composers in gray in Figure 3(b))[17]. Note that nationalist composers are spread all around the South part of the map. This reflects both the idea that Nationalism is the continuation of a 19th century trend and that the network of influences of nationalists might be driven by their citizenship.

Other styles as represented by the Avant-Garde (in saddlebrown) and Experimentalists (in sandybrown) in the Nord-East of the Figure 3(b) pursued the exploration of the 'Uncommon Practice' of the 20th century music. As mentioned by Magnuson (2008), new technology created Electronicism (Varèse, Luening, Babbitt, Maderna, Berio, Stockhausen, Druckman, Davidovsky, Wuorinen) while in the second half of the 20th century there has been an unprecedented attention to new elements of Texturalism – the relationships of timbre, density of pitch and rhythm being given a new primordial status relative to melody and harmony (e.g., Varèse, Carter, Lutoslawski, Ginastera, Ligeti, Xenakis, Stockhausen, Penderecki). Reactions to these styles created

---

[15] Some other composers associated in part with impressionism in Figure 3(b) are Koechlin, Roussel, Schmitt, L. Boulanger, Falla, Mompou, Griffes, Clarke, Carpenter, Karg-Elert, Szymanowski, Malipiero, and Casella.

[16] Besides Stravinsky and Bartok, some composers who have been associated (at least partly) with the neoclassical current are: Hindemith, Prokofiev, Satie, Ibert, Piston, Schulhoff, Martin, Martinu, Orff, Thomson, Tansman, A. Tcherepnin, Chavez, Copland, Weill, Berkeley, Lambert, Tippett, Badings, Holmboe, Francaix, Britten, I. Fine, Persichetti, Foss, Pinkham, Leighton, Harbison, including a group of French composers (Poulenc, Tailleferre, Milhaud, Auric) representing four members of 'Les six' a group of composers often seen as a reaction against both the musical style of Wagner and the impressionist music of Debussy and Ravel.

[17] Composers who have been associated with nationalism at some point of their life are, for example, Vaughan Williams, Holst, Grainger, Rubbra, Kodaly, Ives, Canteloube, Ponce, Turina, Barrios, Butterworth, Villa-Lobos, Moreno Torroba, Moeran, Warlock, Lara, Harris, Revueltas, Rodrigo, Finzi, Guarnieri, Ginastera, Lauro, etc.

Indeterminism/Chance/Aleatory music (Cage, Xenakis, Feldman, Stockhausen), a reaction against the total control that is the basis for integral Serialism. Minimalism (Riley, Reich, Glass, Adams –in cyan in Figure 3(b)) opposed the ideas of atonality itself and reintroduced the vocabulary and sonority of the Common Practice Period. Neo-Romanticism (pre-1950: Walton, Tippett, Shostakovich, Barber, Britten, and post-1950: Rochberg, Henze, Penderecki, Corigliano) opposed these things too, but also represents a complicated relationship between today's composer (and listener) and the music of the past (as opposed to the late Romantic composers, mentioned above, who belong to the Common Practice Period). Popular music, Jazz, exotic influences and the criss-crossing of styles led to Eclecticism—choosing diverse elements from many different sources. This is the essence of the 20th century but certain composers are put in this group as they simply cannot be placed into neat categories due to their originality and individuality (Scriabin, Ives, Ruggles, Cowell, Partch, Messiaen, Hovhaness, Cage, Berio, Crumb, Gorecki, P.M. Davies).[18]

Switching back to Figure 3(a) to emphasize again nationalities, we encounter in the upper east side of the graph composers who have completely changed the rules of the Common Practice Period and created a canon of musical works sometimes considered to be of greater importance from an evolutionary historical point of view. American composers in red (Ives, Cowell, Partch, Cage, Lou Harrison, Wolff, Feldman, Varèse, Riley, Glass, Adams, Reich, Ruggles, Crawford, Wuorinen, Babbitt, Luening, Carter, Rochberg, Session, and Copland). We also see a series of Italian composers in Magenta (Nono, Berio, Maderna, Dallapiccola, Scelsi, Petrassi) proposing a music far away from the music of Italians pictured on the South of the graph. Opposed to the late-romantic German representatives in the East side, we see here a series of decisively 'modern-sounding' German composers (Stockhausen, Rihm, Eisler, Wolpe, Henze, Blacher K.A. Hartmann and Ernst Toch). A few other notable composers on this side of the graph are Birtwistle, Tavener, Nyman, P.M. Davies (English), Messiaen, Jolivet, Xenakis and Boulez (French), Schnittke, Gubaidulina (Russians), Ligeti and Kurtag (Hungarian), Kagel (Argentinian), Takemitsu (Japanese), Dun Tan (Chinese), L. Andriessen (Dutch), Penderecki, Gorecki, and Lutoslawski (Polish).

In Figure 3(d), colour-coding by gender (blue for Men, fuchsia for Women), and by race (gold for African-US/European composers) shows how much white/male-dominant the 20th century composers network remains. All in all, there are 12 women (out of 249 composers) born on or after 1862 in the *CMN* database (Beach, L. Boulanger, Clarke, Crawford, Gubaidulina, Larsen, Monk, Musgrave, Oliveros, Tailleferre, Tower, and Zwilich) and just four women (out of 251) born before 1862 (Chaminade, Hildegard, Mendelssohn-Hensel, and C. Schumann). As for race, the *CMN* has a category African-American/-European composers that includes Samuel Coleridge-Taylor (English), Scott Joplin (American), William Grant Still (American), and George Walker (American), all four born after 1862, and thus included here in the group of '20th century' composers.[19] There are also Latin-American composers in the database (colour coded in the figure in green), such as Leo Brouwer (Afro-Cuban) and Agustin Barrios (partly of Guarani origin form Paraguay). Other Latin-American composers who often used native and folk music

[18] Note that more recent developments are not captured in this graph. For newer developments, we refer the reader to Rutherford-Johnson (2017) who describes the state of music after the fall of the Berlin Wall and discusses how much diverse and fragmented contemporary music has become since 1989.
[19] Louis Moreau Gottschalk is another mixed race composer (having a French Créole mother). He was born before 1862 and therefore is not in Figure 3(d).

of their country are, from Mexico: Carlos Chavez, Silvestre Revueltas, Agustin Lara, and Manuel Ponce; From Brazil: M. Camargo Guarnieri and Heitor Villa-Lobos; from Argentina: Alberto Ginastera and Astor Piazzolla; from Venezuela: Antonio Lauro. Although there are Japanese (Toru Takemitsu) and Chinese (Tan Dun) composers in the database, we do not identify them in Figure 3(d).

The composer selection criteria for inclusion in the *CMN* were objective criteria, not subjective preferences and the *CMN* was not designed as an advocacy instrument. The particular 500 composers selected scored highest on a combination of eleven (unweighted) variables such as length of composer entry in the *Grove's Dictionary of Music* and other catalogs, total number of recordings referring to each composer, and total number of recordings over the past five years only, holdings of sheet music (scores) and other items in 50000 libraries in the U.S and worldwide (through searches in the OCLC *WorldCat* database). See details in the *CMN* website section on Statistics. Hence, the *CMN* aimed at reflecting composer's status at the time it was put together. This said any list of 500 composers (when lists of ten thousands composers exist) will always be open to criticism that 'some other' composers should have been included into the list. There is a relatively new discourse within music departments that the narrative of Western classical music has privileged white men of the 'European' tradition. In so far as the selection criteria above are based on items and catalogues that reflect this narrative, then the *CMN* probably corroborates this bias. Advocacy groups for diversity in music have developed databases reporting the names of thousands of female composers and their compositions (e.g., composerdiversity.com). If these new lists eventually have a large impact on recordings, length of composer entries in dictionaries, etc., then the *CMN* will eventually get out of date as the relative status of composers changes and rapid increase in interest for 'new' or for 'rediscovered' composers arises. In this case an update of the *CMN* will be required. Florence Price (1887-1953), the first African-American woman to be recognized as a symphonic composer and the first to have a composition played by a major orchestra, is perhaps one of those 'rediscovered' composers who could eventually made the list of an updated *CMN*, but so would Francesca Caccini (1587-1640), Barbara Strozzi (1619-1677), Isabella Leonarda (1620-1704), Antonio Cesti (1623-1669), Ferdinand Ries (1784-1838), Louise Farrenc (1804-1875), Allan Pettersson (1911-1980), Mieczyslaw Weinberg (1919-1996), Galina Ustvolskaya (1919- 2006), Alexander Goehr (1932-), Helmut Lachenmann (1935-), Brian Ferneyhough (1943-), Gérard Grisey (1946-1998), Christopher Rouse (1949-2019), Kalevi Aho (1949-), Kaija Saariaho (1952-), Judith Weir (1954-), Magnus Lindberg (1958-), George Benjamin (1960-), Jennifer Higdon (1962-), Eric Whitacre (1970-), Thomas Adès (1971-) and (possibly many) other male and female composers.

As said before, given the criss-crossing of styles, colour-coding nodes/composers according to styles is at best an approximation. A second route would be to use algorithms of 'community detection' and 'cliques'. However, most of these algorithms have been built and used for undirected networks. Instead of pursuing this route, we propose to extend and enrich our analysis by computing similarity indices across composers taking into account both the network of influences of composers and the ecological/musical characteristics that best describe these composers. Section 3 pursues this objective and offers further visualization schemes.

3. Composer similarity indices and heat maps

In order to build similarity indices between composers, two basic sets of information given in the *CMN* have been used.[20] First, we extracted 298 'ecological' categories from the 'Index of Forms and Styles' web page so that each of the 500 composers are associated with a subset of these ecological categories (i.e., characteristics such as time period, geographical location, school association, instrumentation emphases, etc.).[21] An 8-page list of all ecological categories is available in Smith and Georges (2015). Second, we have used the list of 'musical influences' given in the main 'Composers' list page. In particular, for each specific subject composer we extracted 'personal musical' influences (i.e., other composers) who the literature suggests influenced the subject composer.[22] We also extracted 42 more general 'style' influences (e.g., African music, Native American music, Spanish music, Indian music, folk music (by specific regions), popular music (by specific regions), Gypsy music, world music, jazz, ragtime, blues, electronic, gamelan music, nature sounds, birdsong, etc.) also provided in the 'musical influences' list of the *CMN* 'Composers' page.[23]

Once this information was gathered for the 500 main composers of the database, we then constructed bilateral similarity indices based on an approach akin to biosystematic analyzes of biotas or species relations, by means of pairwise comparison of presence-absence data. In essence, we inferred similarities among composers by assuming that if two composers share many of the same musical influences and ecological categories, their music will likely have some similarity. On the other hand, if two composers have very distinct sets of musical influences and ecological categories, then their music is likely to have little similarity.

Technically, for any pair of composers (i,j) for $i,j \in C$ (among the n x n possible pairs with n=500 composers included in the set C of the *CMN* composers page), we have a set A_i of all attributes k (musical, that is personal and style influences, and ecological categories) that apply to composer i, and a set A_j of all attributes k that apply to composer j. We are interested in capturing whether an attribute k applies to both i and j, to i but not j, to j but not i, and to neither i nor j. Thus, for any pair (i,j), $A_i \cap A_j = CA_{i,j}$ is the set of attributes that are related to both i and j; $A_i - A_i \cap A_j = A_{i,-j}$ is the set of attributes that are related to i but not j; $A_j - A_i \cap A_j = A_{j,-i}$ is the set of attributes that are linked to j but not i

[20] The general method described here has been explored and progressively refined in several papers such as Smith and Georges (2014, 2015) and Georges (2017).
[21] See the 'Index of Forms & Styles of music' in the *CMN*. For example, the ecological characteristics associated with Debussy are represented by the following elements: {ballets 1900 on, cello chamber music, chamber music 1825 to 1925, 'Dance' in composition title, etudes, flute unaccompanied, flute chamber music, harp chamber music, harp orchestral music, Impressionist style, nocturnes, operas 1900 on, orchestra incidental music, orchestra symphonic poems, orchestration, Paris composers 1800 on, piano unaccompanied 1775 to 1900, piano unaccompanied 1900 on, piano chamber music general, quartets for strings, song cycles and collections, songs 1800 to 1900, songs 1900 on, suites, trios for other combinations, viola unaccompanied or chamber music, violin chamber music 1850 on}.
[22] See the 'composers' page of the *CMN*. For example, the set of composers who had a documented positive influence on Debussy is: {J.S. Bach, Wagner, Chopin, Tchaikovsky, Liszt, R. Schumann, Ravel, Fauré, Grieg, Rimsky-Korsakov, Mussorgsky, Franck, Gounod, Massenet, Satie, Borodin, Rameau, Albéniz, F. Couperin, Joplin, Delibes, Chausson, Lalo, Chabrier, Dukas, Alkan}.
[23] In the example of Debussy: {Asian music, gamelan music, Renaissance Period music}.

and $DA_{i,j} = A_{i,-j} \cup A_{j,-i}$ is the set of attributes that apply to either *i* or *j* but not both. From this we can produce a table for any pair (*i,j*) that counts the attributes in each of the three sets $CA_{i,j}$, $A_{i,-j}$, $A_{j,-i}$, resulting in corresponding counts, *a*, *b*, and *c*. Given all existing attributes in the database, n_k, we can also count the attributes that belong to neither *i* nor *j* as $n_k - a - b - c = d$. From this we can compute similarity indices for all pairs of composers (*i*, *j*) on the basis of well-known formulas. The centralised cosine similarity measure that we have used is based on earlier literature in scientometrics and bibliographic couplings. The formula is:

(1) $$CSC_{i,j} = (ad - bc)\Big/\sqrt{(a+b)(c+d)(a+c)(b+d)}\ .$$

See Appendix 2 for the derivation of the formula in Eq. (1) and how this is connected to Pearson correlation coefficient, *r*, taken between two Boolean vectors of attributes *k* describing a pair of composers (*i,j*) where the Boolean vector representing a composer *i* is a series (of length n_k) of 1's and 0's when an attribute belongs or not to a composer.

It can be shown that values of the centralised cosine measure range from -1.0 to 1.0. A value of 1.0 indicates that two composers are identical. A value of -1.0 indicates that two composers are complete opposite. A value of 0 shows that two composers are independent (unassociated). A nonzero value of the centralised cosine measure might be due to randomness or actual association between composers. As shown in Smith et al. (2015), unlike in the case of the ordinary cosine measure, there is a proper statistical significance test.[24] Under the assumption that the size of the attribute database n_k is large enough, the distribution of the centralised cosine measure (under the assumption of independence) is approximately normal, with mean 0 and variance $1/n_k$ where n_k is the size of the database at hand, that is the number of attributes *k* characterising all composers in the database.[25] Therefore, the distribution of the centralised cosine measure can be converted into a standard normal distribution using the Z-score/statistics:

(2) $$Z = CSC\Big/\sqrt{1/n_k} \Rightarrow Z = ABS(CSC\sqrt{n_k})\ ,$$

where *ABS* is the absolute value.[26]

This methodology permits us to build a 500x500 matrix of similarity across composers. Let us call this similarity matrix S_{comb} where the subscript '*comb*' refers to the fact that we

---

[24] This test is originally proposed by Giller (2012) who summarizes the statistical properties of statistics computed from independent random bitstreams and derives the moments of the asymptotically normal approximation to the sampling distribution of the cosine similarity of independent random bitstreams. He proposes a new statistic, the support adjusted cosine similarity (where the support is the count of the non-zero bits divided by the length of the bitstream) and notes the parallel between the support adjusted cosine similarity and the Pearson correlation coefficient.

[25] $n_k = 500$ if we just focus on personal musical influences, 542 if we also include more general style influences, 298 if we only include ecological characteristics, and 840 if we include all attributes (all possible music influences and ecological characteristics).

[26] We take the absolute value *ABS*, because square root of 1 is ± 1. Given Eq. 2, the *Z*-statistic is at its critical significant value at 5% when Z = $ABS\left(CSC \times \sqrt{n_k}\right)$ = 1.96. The value for n_k is 840 when all ecological characteristics, styles and personal musical influences are taken into account. Thus, the critical values are $CSC_c = \pm 1.96/\sqrt{840} = \pm 0.067626$. If $CSC_{i,j} > CSC_c = 0.067626$, then *CSC* is considered statistically different from zero so that there is a statistically significant association between composers *i* and *j*.

combined all attributes *k* (musical influences and ecological categories) when computing the similarities across pairs of composers (*i,j*). Figure 4 represents the information obtained from S_{comb} under the form of a heat map.[27] Composers have been classified chronologically, from Hildegard (born in 1098) until Tan (born in 1957). Along the diagonal, composers are compared to themselves. This comparison receives a similarity score of 1 and this translates into a black color code in Figure 4. Moving off the diagonal implies comparing different composers. Dark blue implies high similarity, while a light yellowish color suggests that the two composers are unassociated (independent) and any whiter shading implies negative values (opposition between composers). In general, the further away we move from the diagonal and the more independent the composers' pairs is. We also clearly see intra-period similarities and inter-periods dissimilarities or independence. For example, Renaissance composers tend to be relatively similar (or closer) among themselves, but their music is largely independent from other periods. Note that the database includes many more modern composers than earlier periods' composers, which is translated on the map as seemingly larger 'areas' of darker color for later periods (intra-period similarity).

Of course, we can also build other similarity matrices by restricting attributes to one specific category. For example, we could limit our interest to attributes *k* that focus only on the 298 ecological categories, computing a 500x500 similarity matrix S_{ecol} which provides similarity indices across pairs of composers based on ecological categories only. Or we could limit our interest to attributes *k* that characterise personal musical influences and compute a 500x500 similarity matrix S_{infl}, which would provide similarity indices across pairs of composers based on personal musical influences only. What we have now included in the *CMN* are similarity indices based on the similarity matrix S_{comb}, from which we searched, for each of the 500 subject composers, the top-15 most similar composers. Visually, in Figure 4, it is as if we searched for each composer the 15 darkest composers' pairs. In Table 5 we provide such a list for 20 major composers together with the CSC similarity value/score. The *CMN* now includes this information for all 500 composers. To these scores can be attached statistical significance levels as described above, but for the layperson the scores themselves are easier to appreciate: As now mentioned in the *CMN*, generally speaking, scores above .60 represent composer similarities that are likely to be fairly obvious, scores of about .45 to .60 signify a considerable similarity, .30 to .45 some similarity (for example, of time period or an emphasis on guitar), and below .30 less obvious connections (though many of these may be statistically significant in the greater sense).

It must be re-emphasized that the similarities scores arranged here represent appraisals of correlations between pairings of composers' recorded 'attributes', *k* (i.e., personal and style influences and ecological characteristics). The *CMN* is constructed such that famous composers have many more recorded attributes than lesser-known (or lesser-studied) composers. Hence, the chance of observing 'matches' between famous similar composers ($A_i \cap A_j = CA_{i,j}$) is larger than the chance of observing matches between famous and less-well known, yet similar, composers, which is also larger than the chance of observing matches between two lesser-known similar composers. J.S. Bach, for example, has many more recorded attributes in the *CMN* database than does Johann Ludwig Krebs, a relatively minor figure greatly influenced by Bach. The result is that Bach shows up as the eighth most similar composer in the Krebs entry, but Krebs does not appear in the Bach list of the

[27] The heat map has been produced using Python Seaborn library and dependencies.

15 most similar composers (because of the extent of 'un-matches', that is, attributes that belong to Bach but not to the lesser-studied Krebs). Nevertheless, the lists do pass, at least largely, an eye test.

4. Multidimentional Scaling (MDS), classification algorithms, and hierarchical clustering

Multidimensional Scaling (MDS) is a technique that generates a map displaying the relative positions of a number of objects based on a given set of pairwise distances between these objects. The following example may help to understand the essence of MDS. Given a geographical map of the American continent and a scale, one can compute the aerial distances between cities. If instead the initial data is a set of pairwise distances between North and South American cities, one can attempt to recover the geographical map of the American continent (within about a symmetry and/or rotation). MDS is a methodology that uses algorithms to implement this idea. Although MDS can generate a two-dimensional 'flat map' that could perhaps, or hopefully, be interpreted as latitude and longitude in the geographical example, the technique *per se* can be used to generate more than two dimensions from a given distance matrix. A third dimension here could be interpreted as relative position along the curved surface of the earth.

We can apply the MDS methodology to the 500x500 matrix of pairwise distances across composers, S_{comb}. In this case, the MDS algorithm aims to position each of the 500 composers in an N-dimensional space (i.e., assigning coordinates) such that the initially computed bilateral distances $d_{i,j}$, between composers are preserved as well as possible, according to an optimisation procedure.[28] Choosing N=2 optimizes composers location in a two-dimensional scatterplot. In this case, given the distance $d_{i,j}$ between composers i and j, the algorithm generates the coordinates (x_i, y_i) and (x_j, y_j). The MDS algorithm typically computes coordinates (x,y) so as to minimize a loss function called 'stress', which is a sum of squared errors between the actual distance across any two composers, $d_{i,j}$, and the predicted distance $d_{i,j}$ * computed by the algorithm:

$$\text{Stress} = \sqrt{\sum_{i<j} \left(d_{i,j} - d_{i,j}^{\ *}\right)^2 \Big/ \sum_{i<j} d_{i,j}^{\ 2}} \ ,$$

and where the predicted distances depend on the number of dimensions kept and the algorithm that is used. Stress values near zero are the best.[29,30]

[28] Our initial composers' proximity matrix does not represent pairwise distances across composers, $d_{i,j}$, but pairwise similarities, $s_{i,j}$. Typically similarity indices are converted into distance indices using the formula: $d_{i,j} = \sqrt{s_{i,i} + s_{j,j} - 2s_{i,j}} = \sqrt{2(1 - s_{i,j})}$. As $s_{i,j}$ is given by the *CSC* formula in Eq.(1) that is shown (in Appendix 2) to be equivalent to Pearson coefficient on Boolean vectors, we could alternatively have used the Pearson distance metric (1- $s_{i,}$). In both case $d_{i,j}$ falls between 0 and 2.

[29] There exist several types of MDS algorithms, and they differ mostly in the loss function they use. They are at least two dichotomies that allow to structure some possibilities. 1. Distance scaling (Kruskal-Shepard MDS) versus inner product scaling (classical Torgerson-Gower MDS). 2. Metric scaling (using the actual values of the dissimilarities) versus nonmetric/ordinal scaling (interpreting dissimilarities in terms of the ordination of the data). See Buja et al. (2008) for details. Here we use distance metric scaling.

[30] Following Kruskal (1964), a value of 0 is a perfect goodness-of-fit, 0.05 is good, 0.1 is fair and 0.2 is poor. More recent articles caution against using this advice since acceptable values of stress depends on the quality of the distance matrix and the number of objects in that matrix.

Figure 5 shows the MDS map that we have computed in Python.[31] A general objective of this section is to gauge whether the MDS map places composers according to our general expectations. In order to assess the placement of composers, we tagged composers within 10 periods (Medieval, Renaissance, Baroque, Pre-Classical, Classical, Post-Classical, Early Romantic, Middle Romantic, and late Romantic, and Modern), essentially using their birthdate as a criterion of decision. We then color coded the dot representing each composer on the MDS map for a quick visual check of the placement of composers by periods in which the composer belongs. In essence we see that the map unfolds the history of classical music, starting with Medieval and Renaissance composers in the East, and, as we move counter-clockwise as centuries pass by, progressing towards Baroque, Classical, Romantic and eventually modern composers. Note that the date of birth of composers is never used directly in our methodology to assess the similarity of composers. Of course this does not mean that there is no time dimension in our data set. Clearly, data on the personal musical influences of a subject composer will also include some contemporary composers, and the ecological data have also general references to periods.

Although the first visual check seems to confirm that composers' placement on the map is adequate, we want to produce a more convenient visual check with a painted contour around composers belonging to the same period while producing a map that is esthetically more pleasant than Figure 5. In order to realize this slightly more 'artistic' map in Python, we decided to use classification algorithms such as support vector machines or K-nearest neighbors. Typically, a classification algorithm tries to determine the class to which the object of the analysis belongs to. In the case of music composers, we could have a trained data set of composers and their features (characteristics) from which the algorithm would extract classes. Perhaps to better visualise the approach, suppose that a large number of composers could be represented by just two features and plotted in a two-dimensional graph, then the support vector machine algorithm would try to draw a line between two or more classes of composers in the best possible manner.[32] Then, using a test data set (new composers not included in the trained data set) the algorithm would be used to predict the probability for a new composer to belong to a specific class or group on the basis on his/her features (i.e., his/her positioning on the two-dimensional graph). There are several classification algorithms, both using supervised learning (SVM, K-Nearest Neighbors) and unsupervised learning (K-means clustering). Weiss (2017) and Weiss et al. (2018) apply several methods using audio features to characterise and then classify composers.

Normally, SVM should be applied, in our context, to a set of composers described by a series of features. Here, though, we will apply SVM directly to our MDS map, which characterises composers using two dimensions, the two axes of Figure 5. These two dimensions do not directly represent specific features of composers but instead coordinates ultimately derived from a distance matrix computed on the basis of personal musical influences and ecological/musical features of composers). The reason why we chose this strategy is that our interest is not in predicting to which class a new composer (not included

[31] All MDS maps, in particular Figures 5 and 11, have been generated with Python library Scikit-learn/Manifold. Although Kruskal' value is high at 0.35 for the two-dimension map in Figure 5, suggesting that adding a third dimension might improve the placement, we do not pursue this route here. A '3D' MDS map is shown in the Graphical Appendix (Figure A2).

[32] For SVM, the 'best possible manner' means, in essence, finding a separating line (a hyperplane) between any two groups of composers while producing the widest margin (distance between two parallel lines where each line touches at least a point/dot (a composer) in each class). Support vectors are those points that lie on the two separating margins.

in the trained data set) belongs to. Instead, SVM is used as an algorithm that permits to draw painted contours around composers of a same group and to produce a map that is more 'artistic' than Figure 5 while providing a better or easier visual check of the placement of composers on the MDS map. On a further note, with classification algorithms we typically face a trade-off between fitting the training data set perfectly (high bias so that all composers initially tagged within a same music period will belong to the same class) and how accurately the algorithm can predict the class of a new data set (low variance and consistent predictions using different composers datasets). However, we insist again on the fact that our objective is artistic map drawing with contours for composers classes that perfectly match music periods to which composers have been initially tagged. Hence overfitting is our objective here. Finally, note from Figure 5 that we need to use non-linear SVM because it would be impossible to draw straight contour lines to separate each groups of composers. In other words, we need to bend the lines to separate the classes. For this specific problem we decided to use a non-linear kernel, the radial basis function or Gaussian Kernel.[33]

Figure 6 shows the results of applying non-linear SVM to the MDS map in Figure 5 while forcing overfitting so that we have a perfect matching between 10 classes identified by SVM (and represented with painted contours) and the ten sets of dots of a specific colour, each colour representing a music period wherein composers have been pre-identified.[34] Besides the overfitting (perfect matching), we also see, perhaps more clearly here than in Figure 5, that the MDS map does a good job of positioning composers according to their periods. As Magnuson (2008) mentions, the Common Practice Period (1600-1900) offered a unified view of music with 'macrocosms' to which composers belonged to, such as the Renaissance, Baroque, Classical or Romantic periods representing a somewhat unified view of music rules and practices for rather long periods of time. The MDS map is successful in identifying these macrocosms and the SVM algorithm, by drawing painted contours, provides an easy visual check. In Figure 7 we re-did the same exercise and obtain similar results using K-Nearest neighbor algorithm, overfitting the data by imposing K=1.[35]

However, still according to Magnuson (2008), music of the 20th century is unique in the flow of Western history in its pluralism. As mentioned in Section 2, composers began to explore a more personal and individual approach to musical creation, forming their own microcosms, or small universes. In this perspective, we want to see the results of applying (and overfitting) the SVM algorithm to the data for 20th century composers, assuming that each composer might be categorised by a main style, that is, mutually exclusive categories such as Impressionist, Expressionist, Neo-classical, etc. (as was also done in Figure 3(b)). These categories are given in the legend of Figure 8 which also includes a category 'Before' representing all those composers from earlier periods, who belong without ambiguity to the Common Practice Period.

---

[33] In machine learning, kernels are functions that transform data from non-linear spaces to linear ones.

[34] Figures 6 and Figures 8-14 have been constructed using Scikit-learn/SVM and dependencies. With the radial basis function kernel, two parameters can be chosen: C (the penalty parameter of the error term) and Gamma (which defines how far the influence of a single data observation reaches). A high Gamma means that points closer to the decision boundary have a close reach, that is, the more the algorithm will try to fit the dataset exactly. Here we have set both C = 10 and Gamma =10 to produce overfitting.

[35] Figure 7 has been constructed using Python library Scikit-learn/ KNeighborsClassifier and dependencies.

Unlike our previous results in Figure 6, we now see in Figure 8 that overfitting the data creates a very complex map with many 'islands' of seemingly isolated composers. If we believe that the MDS methodology used here is accurate (as our test above seems to suggest) in positioning any pair of composers according to their bilateral distance, so that composers closer on the map are more similar, then a conclusion can be drawn—Categorizing classical composers of the 20th century is a rather complex task and identifying composers by one main/unique style to generate the contours/regions to which they belong using an overfitting classification algorithm provides little pedagogical guidance in terms of communicating music and trends of the 20th century. Perhaps in this case, avoiding overfitting of the data is a better approach.

Figure 9 shows the results of using a non-linear SVM algorithm while reducing overfitting of the data.[36] As expected, with less over-fitting, we do not observe a perfect matching between painted contours and the sets of dots of a same colour. In other words, in a single class or contour we can see dots of different colors, although one single color tends to dominate from which an appropriate classification is inferred for the whole contour). This map seems to produce a more suited framework to explain classical music and trends of the 20th century. First, observe the three painting contours in pale blue, dark blue and grey/black. From the dominating colour of dots inside these contours we infer that they 'essentially', and respectively, represent Impressionists, Nationalists, and Neo-Romantic composers. Note that these three contours are located closer to the late Romantic composers on the map. This makes sense as the composers adopting these styles have kept unchanged several pillars of the Common Practice Period and therefore have several elements in common with late Romantic composers (see again Magnuson, 2008, for tables demonstrating this). On the other hand, contours painted in yellow (Neo-classical), orange (Expressionists and Serial) and red (Avant-Garde and Experimentalists) are positioned further away from late Romantic composers as they changed most if not all pillars of the Common Practice Period.

Second, we can now attempt to rationalize, in music terms, the fact that composers associated with various colours for dots are in a same painted contour. Many composers of the 20th century did not have a unique style that can easily identify them. On the other hand, the position assigned on the MDS map captures a richer aspect of the complexity of a composer style by taking into account the distance metric of Section 3 (based on personal musical influences and 298 ecological/musical categories which a composer might belong to). A composer who we tagged with a specific style could be in a painted contour that mainly represents composers of another style because the composer effectively composed in both styles. For example, Gerald Finzi (1901-1956) is identified with a brown dot (Nationalist) in Figure 9 but is included in the upward greyish contour that is supposed to represent Neoromantic composers. This might suggest that he is a Nationalist with Neoromantic leanings (either through his personal musical influences or ecological characteristics). And this appears indeed to be the case according to the *CMN* website where 'Nationalist (Neoromantic)' is specified under the style/period category for this composer. Going over each 20th century composer to justify his/her position on the MDS map goes beyond the objective of this paper but musicologists could contribute to this issue by exploring further these results, leading to advances in ways we capture the distance between composers in Section 3. For example, the distance matrix could be built using only the ecological categories while abstracting away from the personal musical influences

[36] In particular we reduce the values for the parameters of the non-linear SVM algorithm to C =7 and Gamma =2.

of a composer (in terms of Section 3, using S_{ecol} instead of S_{comb}). This might improve the accuracy of the MDS map, putting composers who have very similar ecological/music niches even closer on the map. As pursued further in the conclusion, however, this might also reduce the 'endogenous' serendipity that the current map offers in terms of new composer discovery when going from one composer to another one in near vicinity.

Figures A3 to A6 in the graphical appendix explore essentially the same issues as those discussed above, but they do place on the MDS map just 249 composers of the *CMN* database representing the modern (20th century) period (as discussed in Section 2) instead of placing all 500 composers representing several centuries of music. The analysis of these graphs is similar to what we have done in this section and is thus left to the reader.

Finally, we conclude this section with the computation of a dendrogram for modern composers. The agglomerative hierarchical clustering algorithm builds a cluster hierarchy displayed as a tree diagram called a dendrogram. In our case, the input for the algorithm is the 249x249 partition of the 500x500 matrix of composers' similarity S_{comb} so as to focus on modern composers only. Typically, the algorithm applied to the composers' similarity matrix begins with each composer in a separate cluster. In the very first step a two-composer cluster is formed between the two most similar composers. Then, at each successive step, the two clusters that are most similar are joined into a new cluster. Several methods are available to compute distance between clusters of composers (as opposed to the distance between pairs of composers, which is the primary input), such as single linkage, complete linkage, simple average, centroid, median, group average (unweighted pair-group), Ward's minimum variance, etc. Here we used Ward's minimum variance method which minimizes the total within-cluster variance. At each step, the pair of clusters with minimum between-cluster distance are merged.

The dendrogram in Figure 11 is the result of this hierarchical clustering procedure and it identifies five clusters (and several sub-clusters) for the modern period.[37] In Georges and Nguyen (2019) we computed dendrograms for the Baroque, Classical, and Romantic periods and then used the identified clusters and sub-clusters to manually draw them directly on several MDS maps, one for each music period. We do not pursue this at this moment for the modern period, but this could be yet another visual check to gauge whether the composer location on the MDS map in Figure 10 roughly corresponds to clusters and sub-clusters identified with the dendrogram in Figure 11.

5. Conclusion

This paper illustrates different information visualization techniques (data visualization) applied to a classical composers' database. In particular we present composers network graphs, heat maps and multidimensional scaling maps (both obtained from a composer distance matrix), composers' classification maps using support-vector machine and K-Nearest Neighbors algorithms, and dendrograms. All visualizations have been developed using Python programming and libraries. The ultimate objective is to enhance basic music education and interest in classical music by presenting information quickly and clearly, taking advantage of the human visual system's ability to see patterns and trends.

In an age offering either inculcation through lists of 'prescribed' composers and compositions to explore, or music recommendation algorithms that automatically propose

[37] Figure 11 has been constructed with Python library SciPy/cluster.hierarchy and dependencies.

what to listen to next, this paper shows an alternative path that might promote active instead of passive composer and music discovery (as with automatic recommender systems) in a way that is also less restricted than inculcation through prescribed lists. Furthermore, high accuracy (relevance) of automatic music recommender systems tends to generate the same type of music so that people get bored quickly. The much discussed concept of serendipity is the idea of a recommender system that can (pleasantly) surprise a listener. Measuring serendipity is not easy or straightforward. One *cannot simply import it from a Python library* (e.g., sklearn) unlike relevance metrics such as non-discounted cumulative gain (NDCG), mean average precision (MAP), recall, precision, etc. We argue however that our current MDS maps endogenously includes a degree of serendipity: composers that are closer on the map are more similar not only because they share the same ecological/musical niches but also because they share the same personal musical influences. However, as argued in Georges (2017), composers who are similar in their personal musical influences may have nevertheless produced music that sounds different in that they belong to different ecological/musical niches (what is referred to in Georges, 2017, as adaptation or music speciation and evolution). Listening to a composer that is in near vicinity to another better-known composer in the current MDS map may, in that sense, lead to new discoveries with sustained serendipity. Further research could possibly compare current MDS maps with maps computed on the basis of a distance matrix that excludes personal musical influences (i.e., excluding the information from influence networks), gaining relevance in terms of similarity accuracy at the cost of lower serendipity.

Finally, the approach in this paper has a few disadvantages in terms of music discovery. First, a listener who decides to listen to a composer based on his/her location on a MDS map still faces the challenge of discovering his or her important compositions. In this case, we argue that the *CMN* website remains an excellent source of information by proposing a list of important compositions for most of the 500 composers in the database. The overall approach would therefore promote active discovering of composers nurtered through a prescribed list of compositions. Second, although discovering and listening to important compositions of 500 composers may be a lifelong process for most, it remains that there are thousands of composers not included in the *CMN* database. Most notably, the *CMN* does not cover the most recent development in classical music. Without alternative databases covering these newer developments, the method applied in this paper cannot be pursued. In this case, inculcation or recommender algorithms remain the only alternatives for discovering new composers. Rutherford-Johnson (2017) in his *Music after the Fall*, offers a retrospective of modern composition and culture since 1989. The book also compiles several lists of composers and their compositions to listen to. On the other hand much research exists towards building algorithms that can suggest recommendations of new composers and music not included in a pre-existing database, in particular 'semantic labeling' of music that uses artificial intelligence on audio files.[38] This highlights the very

[38] 'Semantic labeling' of music applies artificial intelligence and uses supervised machine learning algorithms to build a model from two types of data, input and output (target) data. Audio features (input data) are first extracted from audio files of specific compositions through an automatic music transcription algorithm, while pre-defined tags or labels (output/target data) are assigned by music experts to the same set of compositions. (Tags' or 'labels' are words that make sense to humans when describing music and are thus helpful when searching or browsing for music.) Hence, the 'training' dataset contains examples of input-output pairs and the algorithm 'learns' relations between audio features and 'tags' or 'labels'. Once the model has been trained, the algorithm (instead of music experts) can assign 'outputs' (tags or labels) to new inputs (audio features of compositions not included in the initial training data set) in what is referred to as a 'test' dataset. This procedure is thus useful for music recommendation of new compositions/composers, or to

relevance of a Music Information Retrieval research agenda based on audio files. Yet, computational analysis of music recordings is still a young field of research and, as mentioned by Weiss et al. (2018), '[e]xtracting score-like information from audio—referred to as *automatic music transcription*—is a complex problem where state-of-the art systems do not show satisfactory performance in most scenarios.' Weiss et al. (2018) also note that the audio processing algorithms needed to extract meaningful audio features are often error-prone and do not reach a high level of specificity regarding human analytical concepts. For example, notes specified by a musical score are hard to extract from a recording. Despite this caveat, automatic music transcription coupled with artificial intelligence and machine learning is both a challenging and exciting research area in terms of discovering new music.

establish similarities between a new composition and earlier compositions included in an initial training data set. For an introduction to semantic labeling of music, see Chapter 4 of Knees and Schedl (2016).

Table 1: Composers sorted by out-degree (top-20)

Name	rank	birth year	death year	country	period	period_agg	degree	out_degree	in_degree	betweenness	eigenvector	# of compositions	importance
Stravinsky	16	1882	1971	Russian-French-American	Modern	Modern	144	123	21	0.030017	0.037016	26	4
Debussy	14	1862	1918	French	Late_Romantic	Romantic	144	118	26	0.040403	0.030889	22	3
Bach, JS	1	1685	1750	German	Baroque	Baroque	129	107	22	0.055674	0.000008	40	4
Wagner	6	1813	1883	German	Early_Romantic	Romantic	128	107	21	0.017367	0.001421	16	3
Beethoven	3	1770	1827	German	Post_Classical	Classical	117	101	16	0.015748	0.000204	48	4
Mozart, WA	2	1756	1791	Austrian	Classical	Classical	115	93	22	0.029102	0.000117	63	5
Liszt	12	1811	1886	Hungarian-French-German	Early_Romantic	Romantic	98	76	22	0.014843	0.001625	21	5
Chopin	10	1810	1849	Polish-French	Early_Romantic	Romantic	91	75	16	0.009932	0.000745	21	3
Schumann, R	13	1810	1856	German	Early_Romantic	Romantic	87	75	12	0.005357	0.000782	26	4
Ravel	20	1875	1937	French	Modern	Modern	97	75	22	0.033069	0.052783	26	2
Schoenberg	34	1874	1951	Austrian-American	Modern	Modern	90	74	16	0.018357	0.020069	18	3
Bartók	25	1881	1945	Hungarian	Modern	Modern	85	66	19	0.018553	0.062088	25	4
Brahms	5	1833	1897	German	Middle_Romantic	Romantic	76	59	17	0.010778	0.000888	37	4
Strauss, R	18	1864	1949	German	Late_Romantic	Romantic	64	52	12	0.00783	0.002044	20	4
Mendelssohn	17	1809	1847	German	Early_Romantic	Romantic	64	47	17	0.005396	0.000435	23	4
Berg	72	1885	1935	Austrian	Modern	Modern	56	44	12	0.003838	0.033721	10	2
Webern	75	1883	1945	Austrian	Modern	Modern	53	41	12	0.007119	0.015274	9	2
Berlioz	29	1803	1869	French	Early_Romantic	Romantic	50	40	10	0.001762	0.000316	14	3
Haydn, J	9	1732	1809	Austrian	Classical	Classical	53	39	14	0.00525	0.000093	34	5
Schubert	4	1797	1828	Austrian	Early_Romantic	Romantic	52	38	14	0.003684	0.000337	35	5

Note to Tables 1-4: Several columns in these tables refer to information obtained from the *CMN,* including year of birth and death of a composer and country/citizenship. *Ranking*: The primary ranking of the *CMN* is based on scores received by composers on a combination of eleven (unweighted) variables such as the length of each composer entry in the *Grove's Dictionary of Music*, the total number of recordings referring to each composer, etc. See Smith (2000). *# of Compositions:* This refers to the number of 'Notable Works' given in the *CMN* for each composer. *Importance:* The *CMN* gives a proxy for the 'Quantity of Work Produced' by a composer. For example, the *CMN* has categories such as: 'immense', 'extensive', 'considerable', 'modest', and 'small'. In Tables 1-4, these labels have been associated with a number such as 5, 4, 3, 2 and 1, respectively. When the *CMN* does not report information for the 'Quantity of Work Produced' by a composer, then the number associated here is 0.

Table 2: Composers sorted by in-degree (top-20)

Name	rank	birth year	death year	country	period	period_agg	degree	out_degree	in_degree	betweenness	eigenvector	# of musical compositions	importance
Debussy	14	1862	1918	French	Late_Romantic	Romantic	144	118	26	0.040403	0.030889	22	3
Bach, JS	1	1685	1750	German	Baroque	Baroque	129	107	22	0.055674	0.000008	40	4
Mozart, WA	2	1756	1791	Austrian	Classical	Classical	115	93	22	0.029102	0.000117	63	5
Liszt	12	1811	1886	Hungarian-French-German	Early_Romantic	Romantic	98	76	22	0.014843	0.001625	21	5
Ravel	20	1875	1937	French	Modern	Modern	97	75	22	0.033069	0.052783	26	2
Shostakovich	27	1906	1975	Russian	Modern	Modern	45	23	22	0.024103	0.101246	19	4
Prokofiev	28	1891	1953	Russian	Modern	Modern	38	16	22	0.005542	0.033041	21	4
Stravinsky	16	1882	1971	Russian-French-American	Modern	Modern	144	123	21	0.030017	0.037016	26	4
Wagner	6	1813	1883	German	Early_Romantic	Romantic	128	107	21	0.017367	0.001421	16	3
Britten	26	1913	1976	British	Modern	Modern	28	8	20	0.015405	0.150192	18	3
Bartók	25	1881	1945	Hungarian	Modern	Modern	85	66	19	0.018553	0.062088	25	4
Ligeti	95	1923	2006	Hungarian-Austrian	Modern	Modern	20	1	19	0.000225	0.173536	8	2
Handel	8	1685	1759	German-British	Baroque	Baroque	51	33	18	0.022522	0	23	5
Messiaen	63	1908	1992	French	Modern	Modern	40	22	18	0.004534	0.07876	11	3
Elgar	45	1857	1934	British	Late_Romantic	Romantic	30	12	18	0.002544	0.003239	13	4
Brahms	5	1833	1897	German	Middle_Romantic	Romantic	76	59	17	0.010778	0.000888	37	4
Mendelssohn	17	1809	1847	German	Early_Romantic	Romantic	64	47	17	0.005396	0.000435	23	4
Penderecki	147	1933	2020	Polish	Modern	Modern	21	4	17	0.000602	0.197536	6	2
Crumb	163	1929	--	American	Modern	Modern	19	2	17	0.000249	0.13741	5	2
Ginastera	120	1916	1983	Argentinian	Modern	Modern	18	1	17	0.000232	0.197447	8	3

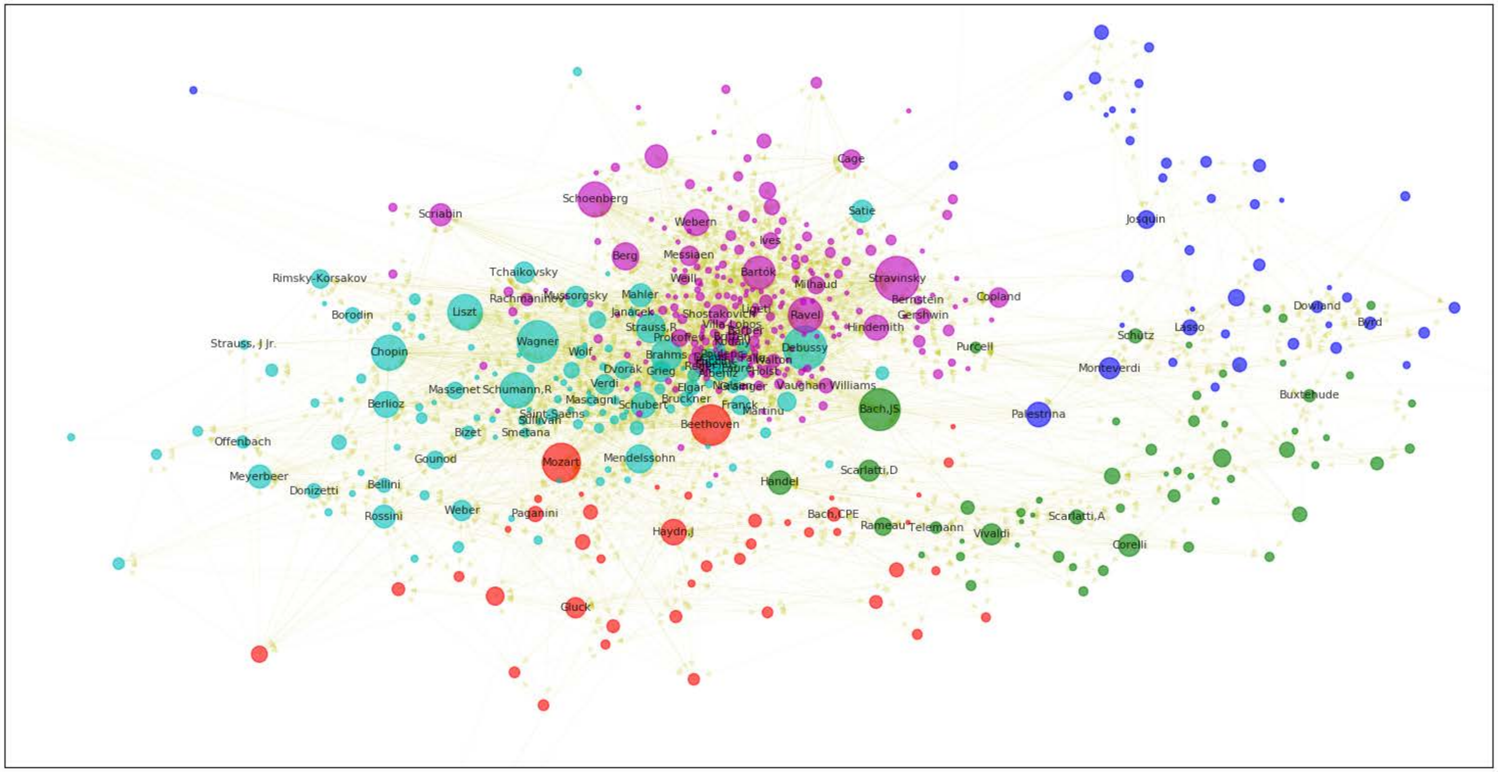
Table 3: Composers sorted by betweenness centrality (top-20)

Name	rank	birth year	death year	country	period	period_agg	degree	out_degree	in_degree	betweenness	eigenvector	# of compositions	importance
Bach, JS	1	1685	1750	German	Baroque	Baroque	129	107	22	0.055674	0.000008	40	4
Debussy	14	1862	1918	French	Late_Romantic	Romantic	144	118	26	0.040403	0.030889	22	3
Ravel	20	1875	1937	French	Modern	Modern	97	75	22	0.033069	0.052783	26	2
Stravinsky	16	1882	1971	Russian-French-American	Modern	Modern	144	123	21	0.030017	0.037016	26	4
Mozart, WA	2	1756	1791	Austrian	Classical	Classical	115	93	22	0.029102	0.000117	63	5
Gershwin	30	1898	1937	American	Modern	Modern	20	13	7	0.026019	0.066307	8	3
Shostakovich	27	1906	1975	Russian	Modern	Modern	45	23	22	0.024103	0.101246	19	4
Handel	8	1685	1759	German-British	Baroque	Baroque	51	33	18	0.022522	0	23	5
Bartók	25	1881	1945	Hungarian	Modern	Modern	85	66	19	0.018553	0.062088	25	4
Schoenberg	34	1874	1951	Austrian-American	Modern	Modern	90	74	16	0.018357	0.020069	18	3
Wagner	6	1813	1883	German	Early_Romantic	Romantic	128	107	21	0.017367	0.001421	16	3
Sweelinck	192	1562	1621	Dutch	Renaissance	Renaissance	14	6	8	0.016239	0	0	4
Beethoven	3	1770	1827	German	Post_Classical	Classical	117	101	16	0.015748	0.000204	48	4
Britten	26	1913	1976	British	Modern	Modern	28	8	20	0.015405	0.150192	18	3
Liszt	12	1811	1886	Hungarian-French-German	Early_Romantic	Romantic	98	76	22	0.014843	0.001625	21	5
Palestrina	80	1525	1594	Italian	Renaissance	Renaissance	41	36	5	0.012841	0	9	4
Brahms	5	1833	1897	German	Middle_Romantic	Romantic	76	59	17	0.010778	0.000888	37	4
Purcell	40	1659	1695	British	Baroque	Baroque	19	6	13	0.010048	0	11	4
Chopin	10	1810	1849	Polish-French	Early_Romantic	Romantic	91	75	16	0.009932	0.000745	21	3
Schütz	89	1585	1672	German	Baroque	Baroque	20	11	9	0.008633	0	5	5

Table 4: Composers sorted by eigenvector centrality (top-20)

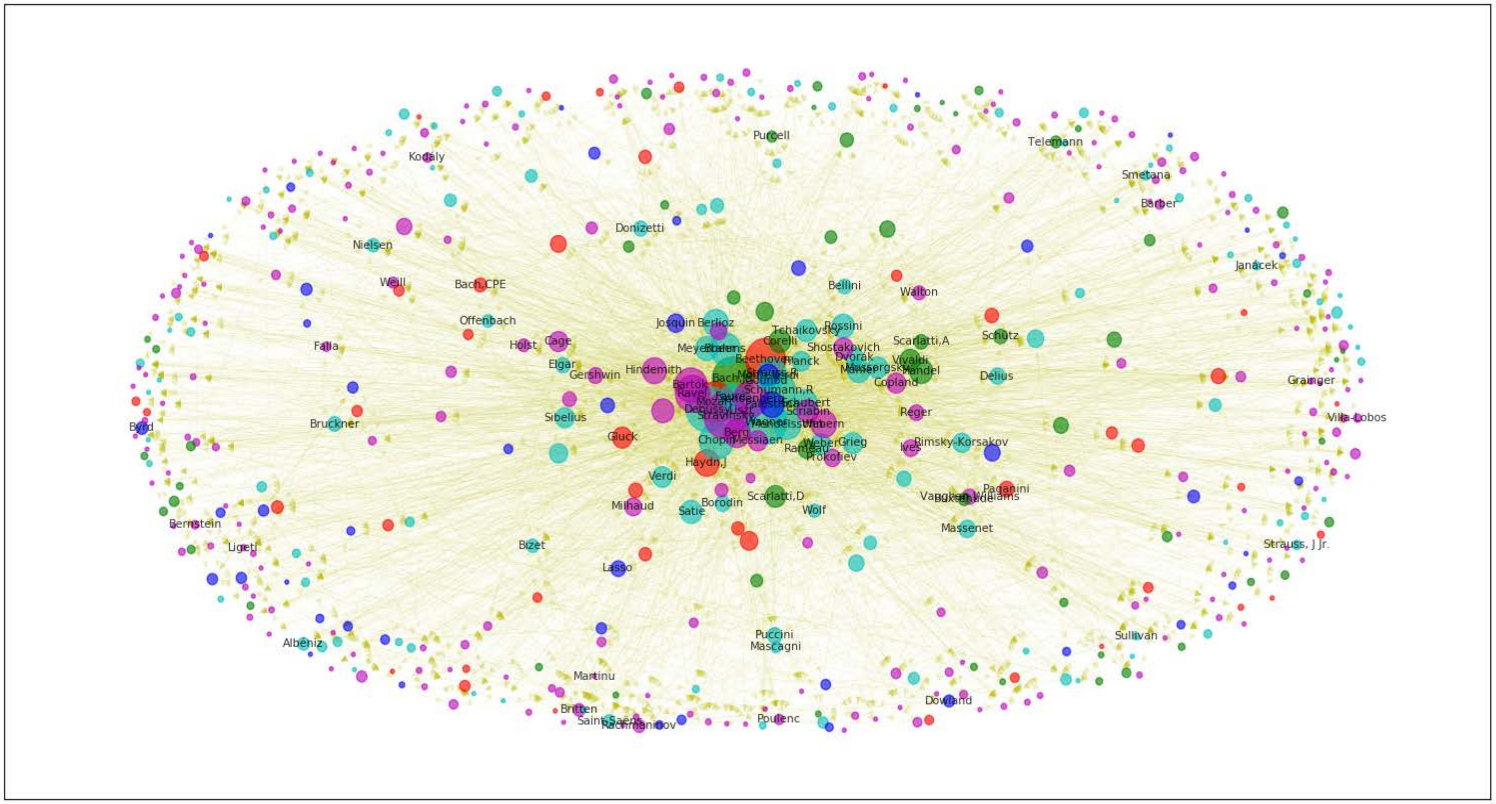
Name	rank	birth year	death year	country	period	period_agg	degree	out_degree	in_degree	betweenness	eigenvector	# of musical compositions	importance
Górecki	287	1933	2010	Polish	Modern	Modern	14	0	14	0	0.201387	3	0
Penderecki	147	1933	2020	Polish	Modern	Modern	21	4	17	0.000602	0.197536	6	2
Ginastera	120	1916	1983	Argentinian	Modern	Modern	18	1	17	0.000232	0.197447	8	3
Brouwer	281	1939	--	Cuban	Modern	Modern	9	0	9	0	0.181309	3	0
Andriessen	360	1939	--	Dutch	Modern	Modern	12	0	12	0	0.17914	0	0
Tan	454	1957	--	Chinese-American	Modern	Modern	11	0	11	0	0.178059	1	0
Ligeti	95	1923	2006	Hungarian-Austrian	Modern	Modern	20	1	19	0.000225	0.173536	8	2
Britten	26	1913	1976	British	Modern	Modern	28	8	20	0.015405	0.150192	18	3
Bennett, RR	253	1936	2012	British	Modern	Modern	10	0	10	0	0.149939	0	0
Kurtág	341	1926	--	Romanian-Hungarian	Modern	Modern	15	0	15	0	0.143342	1	0
Crumb	163	1929	--	American	Modern	Modern	19	2	17	0.000249	0.13741	5	2
Boulez	124	1925	2016	French	Modern	Modern	25	14	11	0.002293	0.13415	5	1
Corigliano	230	1938	--	American	Modern	Modern	10	0	10	0	0.129607	3	0
Cage	59	1912	1992	American	Modern	Modern	35	22	13	0.004594	0.12613	6	3
Adams	191	1947	--	American	Modern	Modern	13	0	13	0	0.12326	6	2
Birtwistle	244	1934	--	British	Modern	Modern	9	0	9	0	0.122441	0	0
Bolcom	183	1938	--	American	Modern	Modern	9	0	9	0	0.122106	0	3
Takemitsu	126	1930	1996	Japanese	Modern	Modern	10	1	9	0.000052	0.118251	4	3
Carter	110	1908	2012	American	Modern	Modern	18	2	16	0.000385	0.11761	8	3
Nyman	311	1944	--	British	Modern	Modern	11	0	11	0	0.115754	2	0

Figure 1: Composers' influence network. ForceAtlas2 algorithm



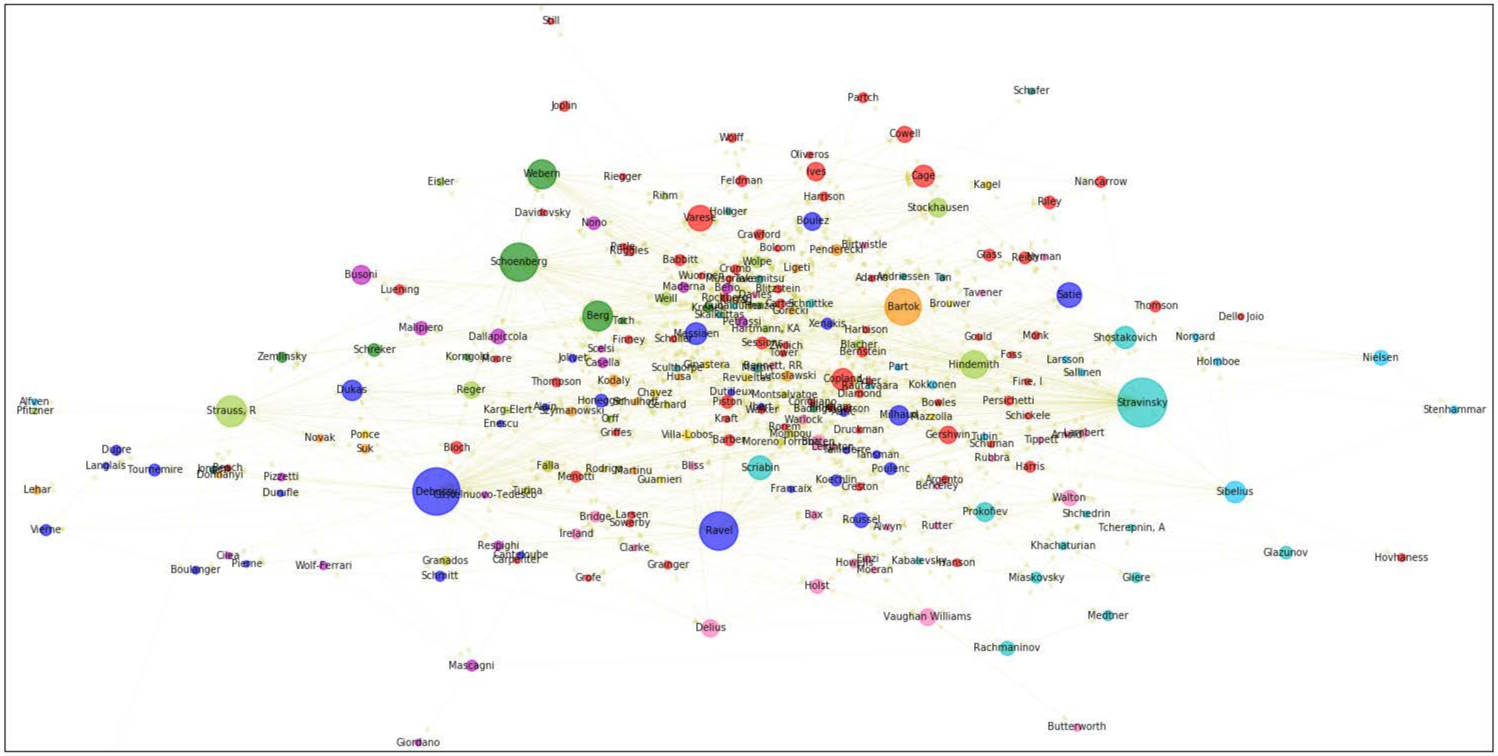
Note: Blue = Medieval and Renaissance; Green = Baroque; Red = Classical; Cyan = Romantic; Magenta = 20th century

Figure 2: Composers' influence network. The Fruchterman and Rheingold algorithm



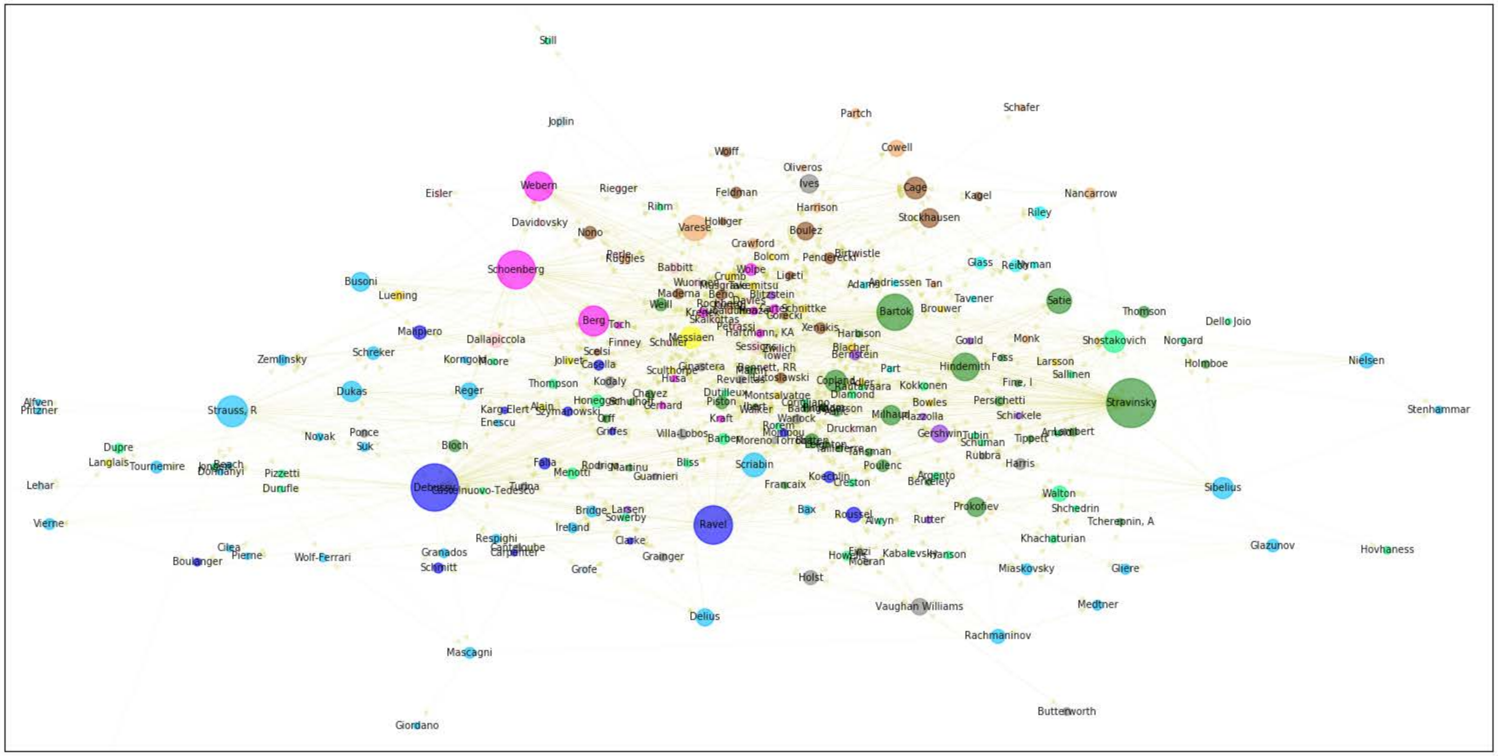
Note: Blue = Medieval and Renaissance; Green = Baroque; Red = Classical; Cyan = Romantic; Magenta = 20th century

Figure 3(a): 20th Century composers' network. Colour code by citizenship. ForceAtlas2 algorithm



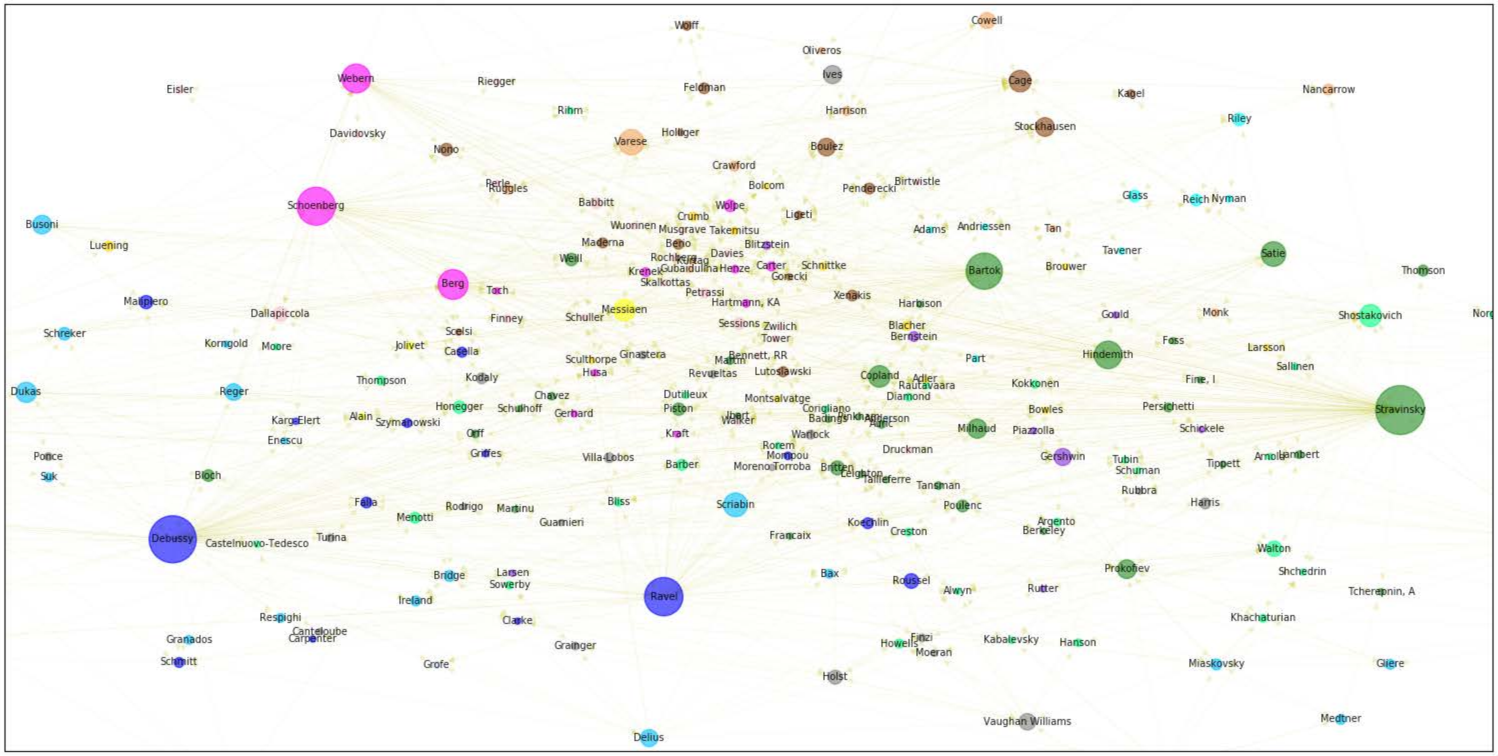
Note: 'Red': Americans; 'Hotpink': British; 'Green': Austrians; 'Yellowgreen': Germans; 'Blue': French; 'Deepskyblue': North-Europeans; 'Cyan': Russians; 'Magenta': Italians; 'Gold': South-Americans; 'Yellow': Spanish; 'Darkorange': Central-Europeans; 'Teal': Other Nationalities.

Figure 3(b): 20th Century composers' network. Colour code by style. ForceAtlas2 algorithm



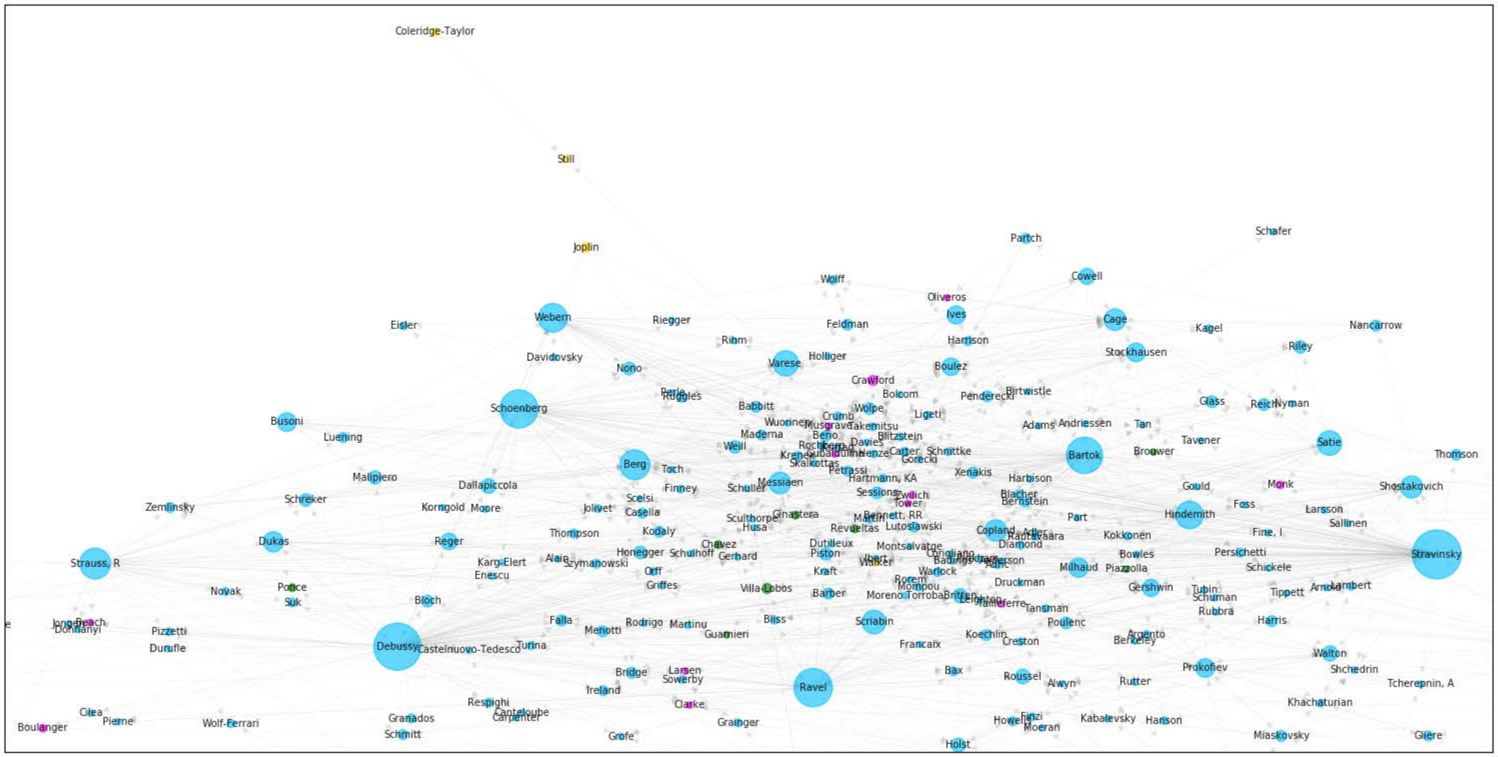
Note: Late Romantic: deepskyblue; Light Classical: lightblue; Impressionist: blue; Nationalist: gray; Vernacularist: blueviolet; Expressionist: magenta; Serial: pink; Neoclassical: forestgreen; Avant-Garde: saddlebrown; Experimentalist: sandybrown; Mystical: yellow; Eclectic: gold; Neoromantic: springgreen; Minimalist: cyan.

Figure 3(c): 20th Century composers' network. Colour code by style. Zoom level 1. ForceAtlas2 algorithm



Note: Late Romantic: deepskyblue; Light Classical: lightblue; Impressionist: blue; Nationalist: gray; Vernacularist: blueviolet; Expressionist: magenta; Serial: pink; Neoclassical: forestgreen; Avant-Garde: saddlebrown; Experimentalist: sandybrown; Mystical: yellow; Eclectic: gold; Neoromantic: springgreen; Minimalist: cyan.

Figure 3(d): 20th Century composers' network. Colour code by gender and race. ForceAtlas2 algorithm



Note 1: Blue for men, fuchsia for women, gold for African-American and African-European composers, green for some Latin-American composers.
Women: L. Boulanger, Beach, Clarke, Larsen, Tailleferre, Tower, Zwilich, Gubaidulina, Musgrave, Crawford, Oliveros, and Monk.
African-American/-European composers: Samuel Coleridge-Taylor (English), William Grant Still, Scott Joplin, and George Walker.
Latin-American: Leo Brouwer (Afro-Cuban); Agustin Barrios (Paraguay); Carlos Chavez, Silvestre Revueltas, Agustin Lara, and Manuel Ponce (Mexico); M. Camargo Guarnieri and Heitor Villa-Lobos (Brazil); Alberto Ginastera and Astor Piazzolla (Argentina); Antonio Lauro (Venezuela).
Note 2: Lauro, Lara, and Barrios are too far off the center of the map to be shown here.

Figure 4: Heat map for 500 composers

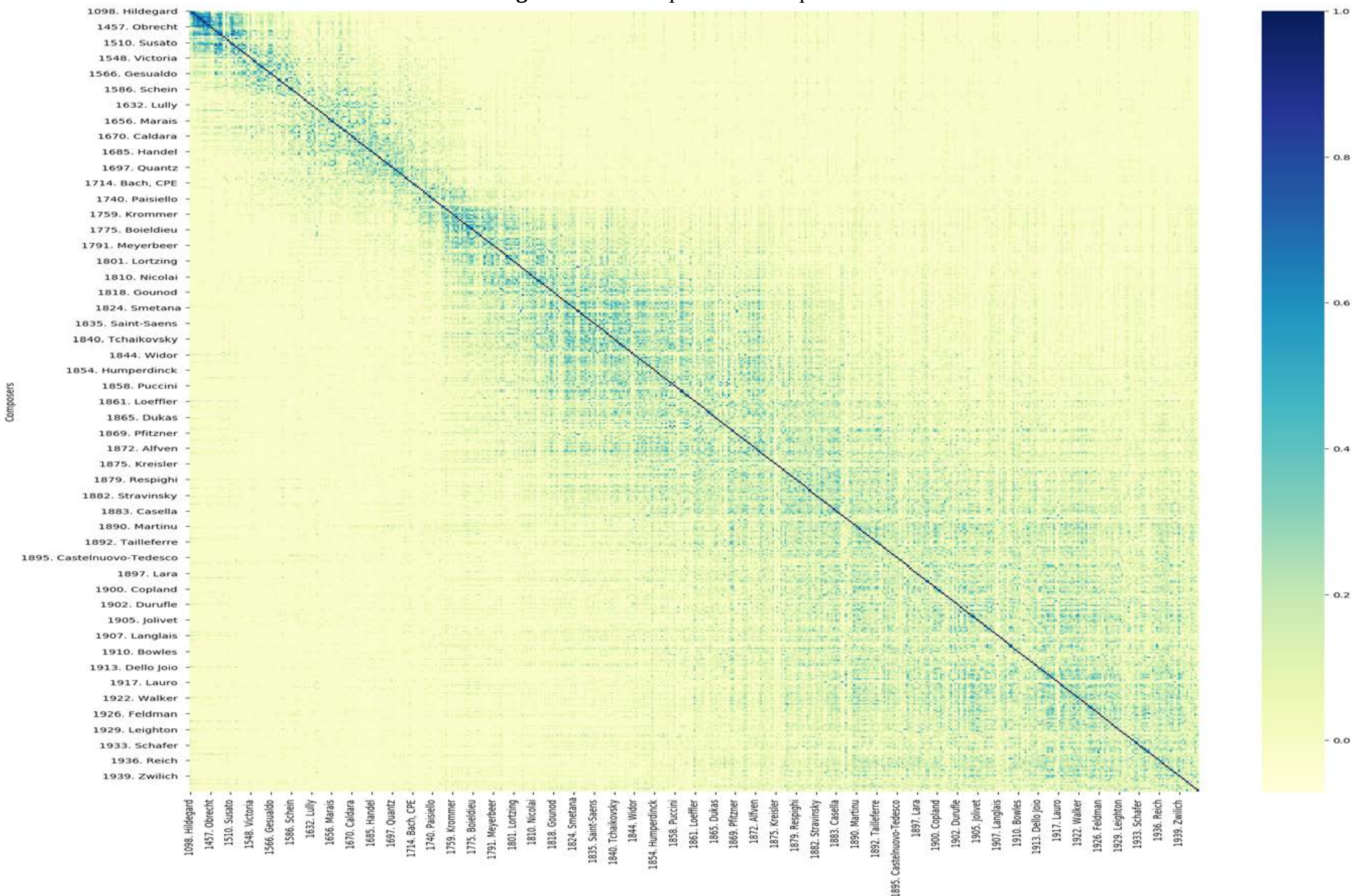


Table 5: Most similar composers to the top 20 composers* (after the *CMN* site)

COMPOSER	*15 COMPOSERS MOST SIMILAR TO*
1. JS Bach	Buxtehude .43; Pachelbel .43; Handel .42; Telemann .38; Böhm .37; F Couperin .34; Vivaldi .33; Biber .33; Bruhns .32; CPE Bach .32; Weiss .32; Scheidt .31; Leclair .31; Muffat .30; Fux .30
2. Mozart	J Haydn .46; Beethoven .39; JC Bach .38; Cimarosa .32; Salieri .32; Boccherini .29; Dittersdorf .29; Gluck .29; Schubert .27; Weber .26; Cherubini .24; CPE Bach .22; Paisiello .22; M Haydn .21; Brahms .20
3. Beethoven	Schubert .46; Hummel .45; J Haydn .43; Mendelssohn .43; Mozart .39; Dussek .34; Reicha .33; Brahms .33; Weber .32; Rossini .31; Spohr .31; M Haydn .30; Cherubini .30; Clementi .29; Field .29
4. Schubert	Rossini .47; Beethoven .46; Mendelssohn .43; Spohr .38; Berlioz .36; Reicha .33; Liszt .33; Weber .32; Carulli .31; Méhul .30; Hummel .30; Kuhlau .29; Giuliani .29; R Schumann .28; Mayr .27
5. Brahms	Dvorák .49; R Schumann .42; Bruch .41; Liszt .40; Fauré .39; Mendelssohn .39; Elgar .38; C Schumann .35; Grieg .34; Franck .34; Mahler .34; Goldmark .33; Rheinberger .33; Beethoven .33; Franz .32
6. Wagner	Nicolai .50; Smetana .45; Gounod .44; Meyerbeer .44; Verdi .43; Glinka .43; Berwald .42; Thomas .38; Donizetti .38; Berlioz *.37; Goldmark .36; Alkan .36; Lortzing .36; Boieldieu .34; Weber .34*
7. Verdi	Gounod .52; Donizetti .46; Wagner .43; Nicolai .41; Berlioz .40; Bruckner .36; Bellini .36; Boito .36; Meyerbeer .35; Mercadante .33; Glinka .33; Offenbach .32; Liszt .31; Elgar .31; Ponchielli .31
8. Handel	Vivaldi .46; JS Bach .42; Purcell .39; Telemann .36; Blow .36; Albinoni .33; Stradella .31; Geminiani .31; Zelenka .30; Bononcini .30; Pergolesi .29; Biber .29; A Scarlatti .28; Rameau .28; Leclair .28
9. J Haydn	Mozart .46; Beethoven .43; M Haydn .37; Dittersdorf .32; JC Bach .30; Boccherini .26; CPE Bach .26; Dussek .25; Hummel .24; Cimarosa .23; Handel .22; Clementi .21; Gluck .20; Schubert .20; Zelenka .19
10. Chopin	Alkan .46; Liszt .42; Field .39; Czerny .38; R Schumann .37; Wagner .32; Berwald .32; Glinka .31; Meyerbeer .29; Nicolai .29; Thomas .29; Busoni .29; Franck .28; Mendelssohn-Hensel .28; Mendelssohn .27
11. Tchaikovsky	Balakirev .54; Borodin .51; Rimsky-Korsakov .50; Rubinstein .46; Mussorgsky .46; Cui .41; Saint-Saëns .39; Arensky .38; Smetana .37; Elgar .36; Dvorák .36; Chabrier .33; Massenet .33; Raff .32; Dargomïzhsky .32
12. Liszt	Franck .48; Chopin .42; Mendelssohn .42; Alkan .42; Mahler .41; Brahms .40; Gounod .38; R Schumann .37; Raff .36; Wolf .35; Berlioz .35; Rheinberger .34; Balakirev .34; Bruckner .34; Reger .34
13. R Schumann	C Schumann .43; Brahms .42; Mendelssohn .40; Mendelssohn-Hensel .38; Liszt .37; Chopin .37; Gade .34; Berwald .34; Fauré .33; Alkan .33; Rubinstein .31; Raff .29; Mahler .29; Schubert .28; Bruch .27
14. Debussy	Ravel .52; Granados .41; Roussel .39; Fauré .39; Koechlin .39; Duparc .35; Chausson .33; Dukas .33; Schmitt .33; Falla .33; Chaminade .31; Rachmaninov .30; Borodin .30; Prokofiev .29; Ibert .29
15. Puccini	Leoncavallo .69; Giordano .64; Mascagni .63; Boito .48; Cilea .45; Ponchielli .43; G Charpentier .40; Wolf-Ferrari .37; Malipiero .30; Massenet .30; Mendelssohn-Hensel .29; Smetana .28; Griffes .27; Pizzetti .27; Bizet .27
16. Stravinsky	Honegger .39; Prokofiev .37; Ravel .34; Shostakovich .33; Berg .31; Hindemith .31; Mompou .29; Glière .29; Dallapiccola .29; Harris .28; Kabalevsky .28; Varèse .28; Tippett .28; Janácek .28; Schoenberg .27
17. Mendelssohn	Beethoven .43; Schubert .43; Liszt .42; R Schumann .40; Brahms .39; Spohr .36; Berlioz .35; Berwald .34; Mendelssohn-Hensel .33; Rheinberger .33; Alkan .32; C Schumann .32; Loewe .31; Field .30; Clementi .30

18. R Strauss Mahler .46; Dvorák .42; Raff .37; Pfitzner .37; Schoenberg .37; Stenhammar .34; Reinecke .34; Wolf .33; Zemlinsky .31; Elgar .31; Cui .31; Reger .30; Smetana .30; Sinding .30; Grieg .29

19. Mahler Wolf .63; Goldmark .46; R Strauss .46; Parry .44; Stenhammar .43; Gade .43; Raff .42; Liszt .41; Schoenberg .40; Stanford .39; Reger .39; Bruckner .38; Rheinberger .38; Pfitzner .37; Franz .36

20. Ravel Debussy .52; Fauré .46; Prokofiev .40; Janácek .39; Poulenc .38; Schmitt .36; Koechlin .35; Moszkowski .35; Falla .35; Ibert .35; Granados .34; Roussel .34; Stravinsky .34; Chausson .33; Milhaud .33

***Note:** The number following each individual composer in each of the top-15 most similar lists is the centralised cosine similarity score obtained from Eq. (1).

Figure 5: Multidimensional scaling analysis (MDS) using bilateral distances between composers

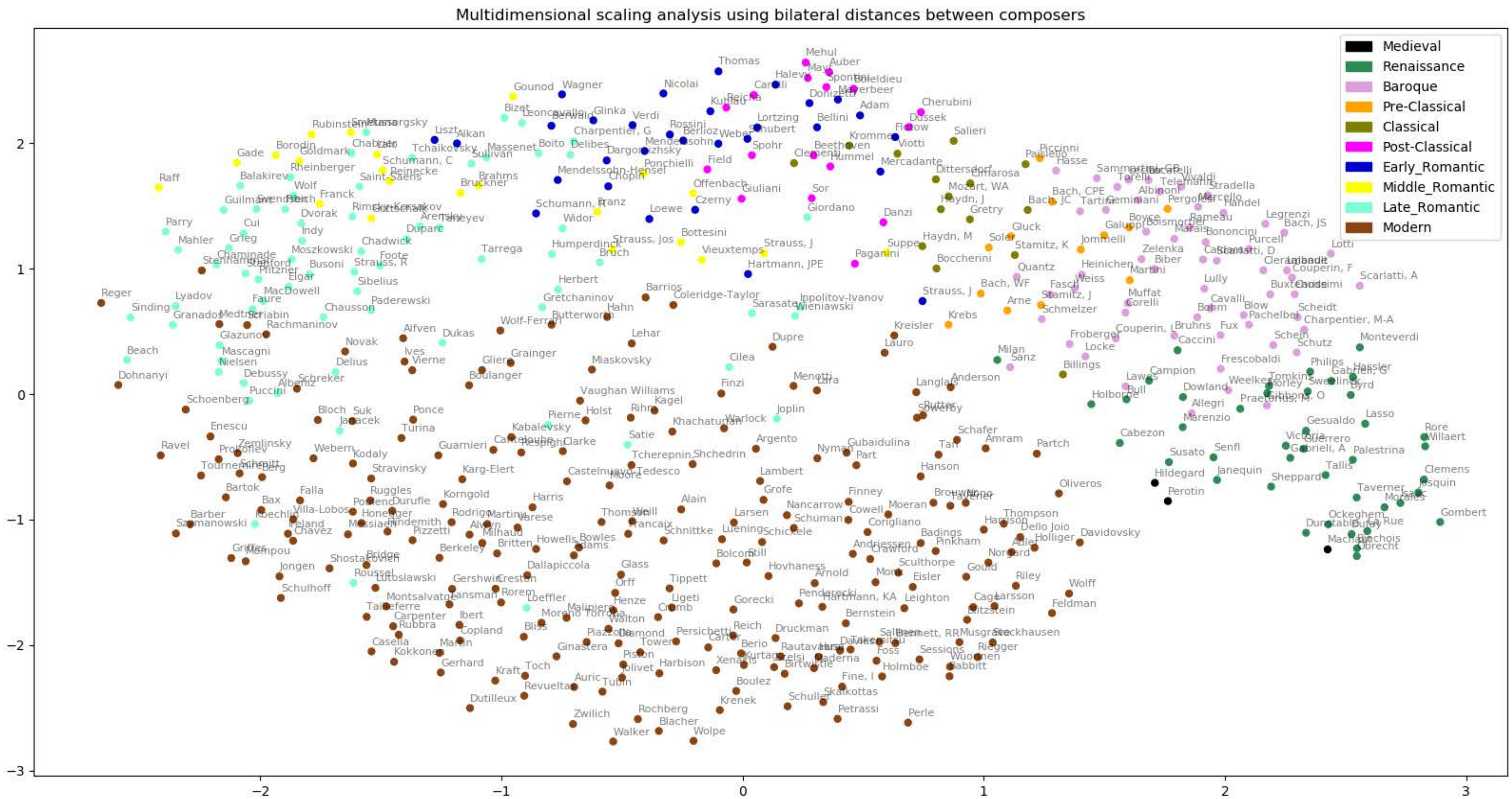


Figure 6: Support vector machines on MDS map — 500 composers with emphasis on the 'Macrocosm' of the Common Practice Period

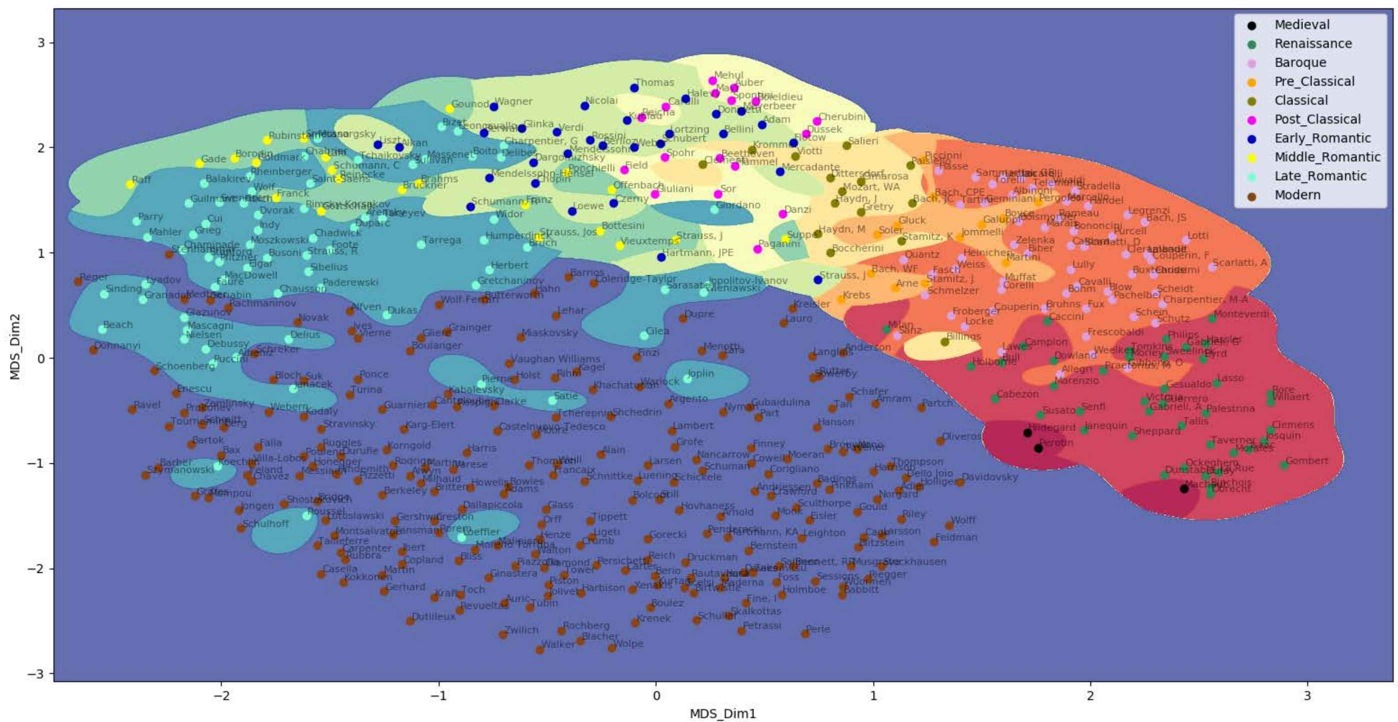


Figure 7: K-Nearest Neighbors on MDS map — 500 composers with emphasis on the ‘Macrocosm’ of the Common Practice Period

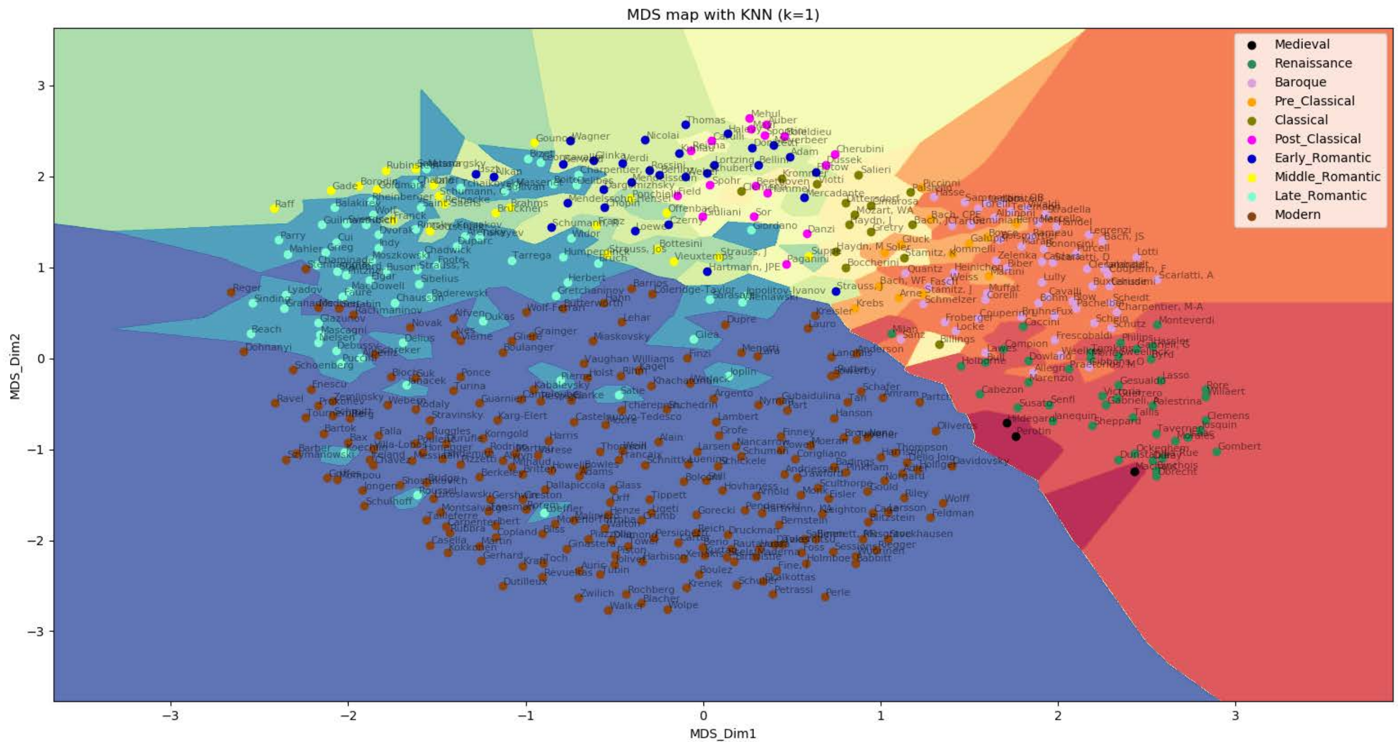


Figure 8: Support vector machines on MDS map — 500 composers with emphasis on 20th century composers and overfitting—The 'microcosms' of 20th century classical music

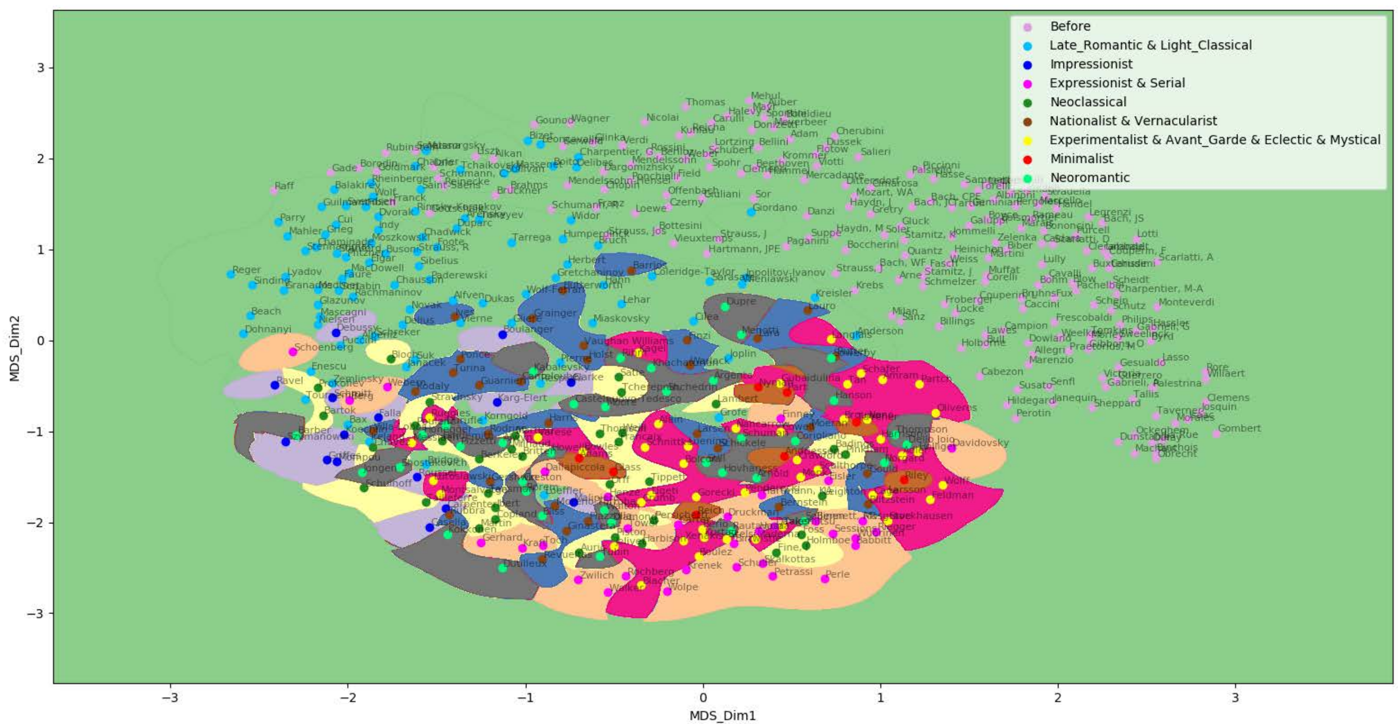


Figure 9: SVM on MDS map — 500 composers with emphasis on 20th century composers and less overfitting

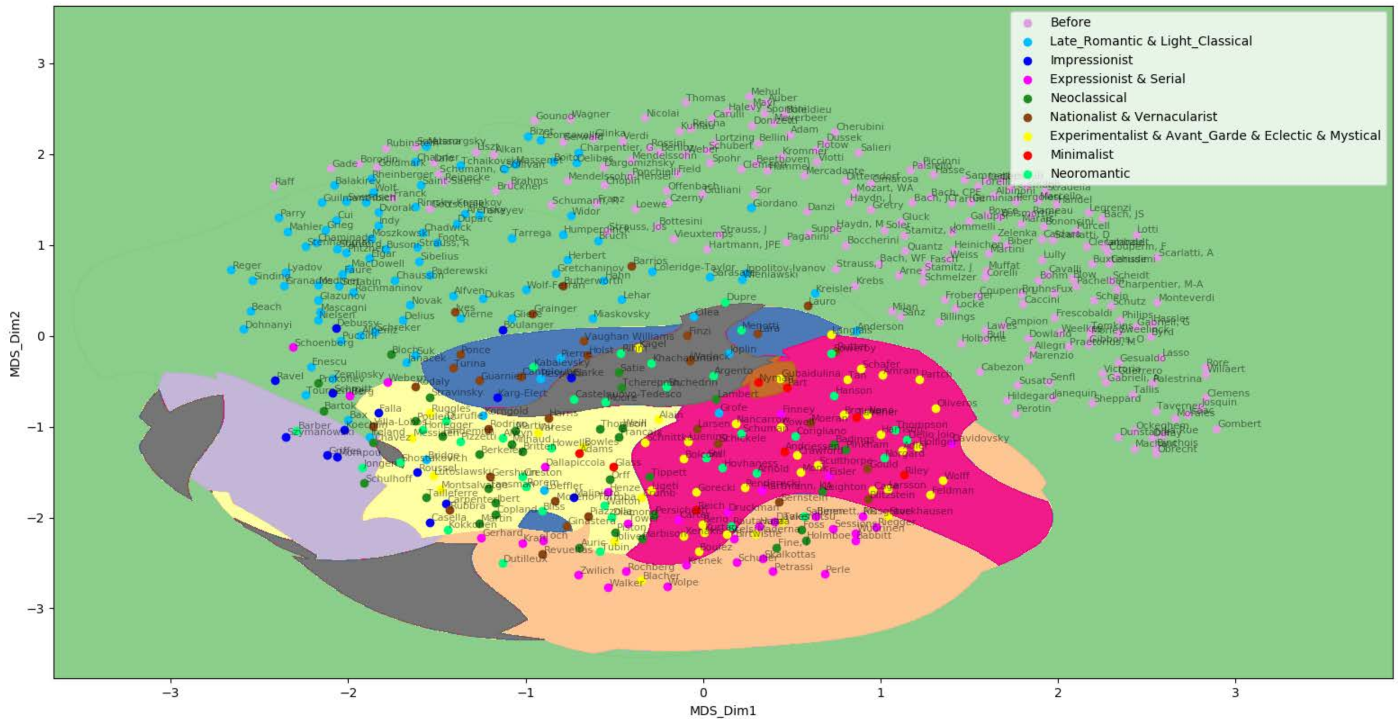


Figure 10: SVM on MDS map — 500 composers with emphasis on 20th century composers – The 'in between' case

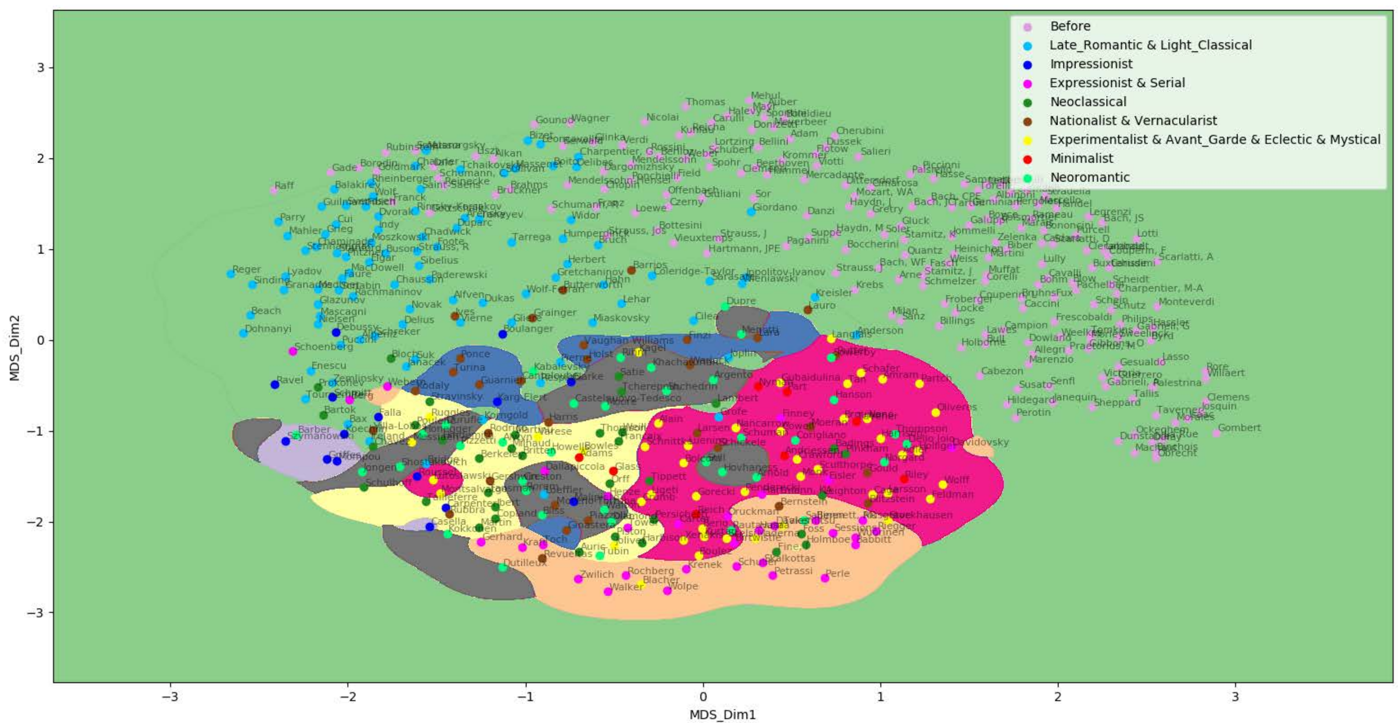
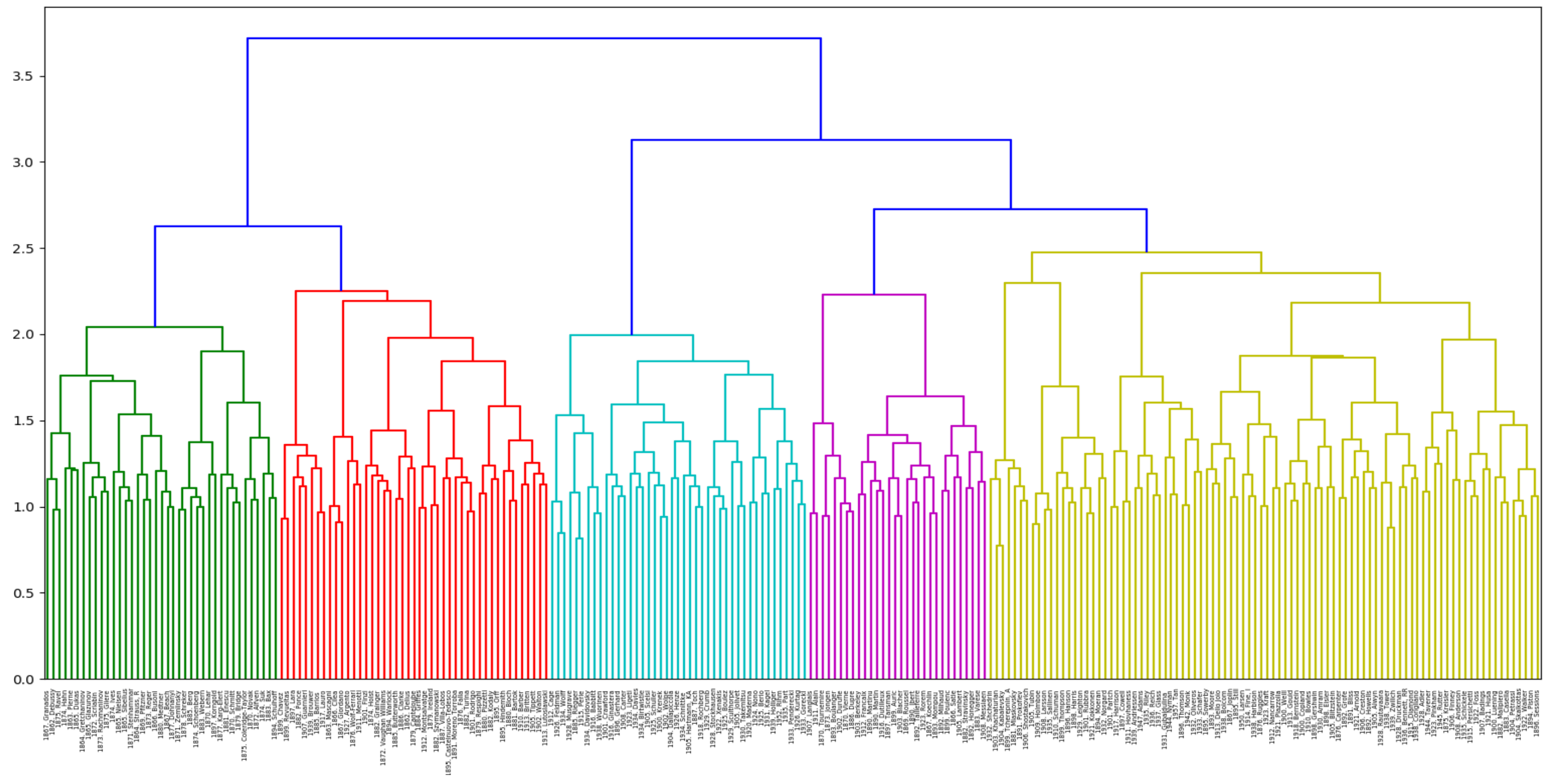


Figure 11: Dendrogram for 20th century composers



Appendix. Shortest paths from any composers to Debussy

1 step (direct influences on Debussy)

[' Bach, JS', ' Debussy']
[' Wagner', ' Debussy']
[' Chopin', ' Debussy']
[' Tchaikovsky', ' Debussy']
[' Liszt', ' Debussy']
[' Schumann, R', ' Debussy']
[' Ravel', ' Debussy']
[' Faure', ' Debussy']
[' Grieg', ' Debussy']
[' Rimsky-Korsakov', ' Debussy']
[' Mussorgsky', ' Debussy']
[' Franck', ' Debussy']
[' Gounod', ' Debussy']
[' Massenet', ' Debussy']
[' Satie', ' Debussy']
[' Borodin', ' Debussy']
[' Rameau', ' Debussy']
[' Albeniz', ' Debussy']
[' Couperin, F', ' Debussy']
[' Joplin', ' Debussy']
[' Delibes', ' Debussy']
[' Chausson', ' Debussy']
[' Lalo', ' Debussy']
[' Chabrier', ' Debussy']
[' Dukas', ' Debussy']
[' Alkan', ' Debussy']

2 steps

[' Mozart, WA', ' Wagner', ' Debussy']
[' Beethoven', ' Wagner', ' Debussy']
[' Schubert', ' Liszt', ' Debussy']
[' Brahms', ' Chausson', ' Debussy']
[' Verdi', ' Mussorgsky', ' Debussy']
[' Handel', ' Bach, JS', ' Debussy']
[' Puccini', ' Ravel', ' Debussy']
[' Stravinsky', ' Ravel', ' Debussy']
[' Mendelssohn', ' Wagner', ' Debussy']
[' Strauss, R', ' Ravel', ' Debussy']
[' Vivaldi', ' Bach, JS', ' Debussy']
[' Rossini', ' Wagner', ' Debussy']
[' Berlioz', ' Wagner', ' Debussy']
[' Gershwin', ' Ravel', ' Debussy']
[' Schoenberg', ' Ravel', ' Debussy']
[' Telemann', ' Bach, JS', ' Debussy']
[' Bizet', ' Tchaikovsky', ' Debussy']
[' Donizetti', ' Chopin', ' Debussy']
[' Saint-Saens', ' Ravel', ' Debussy']
[' Weber', ' Wagner', ' Debussy']
[' Bellini', ' Wagner', ' Debussy']
[' Scarlatti, D', ' Bach, JS', ' Debussy']
[' Gluck', ' Wagner', ' Debussy']
[' Paganini', ' Chopin', ' Debussy']
[' Palestrina', ' Bach, JS', ' Debussy']
[' Schutz', ' Bach, JS', ' Debussy']
[' Corelli', ' Bach, JS', ' Debussy']
[' Meyerbeer', ' Wagner', ' Debussy']
[' Buxtehude', ' Bach, JS', ' Debussy']
[' Glinka', ' Tchaikovsky', ' Debussy']
[' Pachelbel', ' Bach, JS', ' Debussy']
[' Lully', ' Bach, JS', ' Debussy']
[' Charpentier, M-A', ' Couperin, F', ' Debussy']
[' Frescobaldi', ' Bach, JS', ' Debussy']
[' Hummel', ' Chopin', ' Debussy']
[' Albinoni', ' Bach, JS', ' Debussy']
[' Schumann, C', ' Schumann, R', ' Debussy']
[' Spohr', ' Wagner', ' Debussy']
[' Sarasate', ' Albeniz', ' Debussy']

[' Clementi', ' Chopin', ' Debussy']
[' Cherubini', ' Wagner', ' Debussy']
[' Rubinstein', ' Tchaikovsky', ' Debussy']
[' Gottschalk', ' Joplin', ' Debussy']
[' Roussel', ' Satie', ' Debussy']
[' Sweelinck', ' Bach, JS', ' Debussy']
[' Balakirev', ' Tchaikovsky', ' Debussy']
[' Thomas', ' Massenet', ' Debussy']
[' Marais', ' Couperin, F', ' Debussy']
[' Mendelssohn-Hensel', ' Gounod', ' Debussy']
[' Salieri', ' Liszt', ' Debussy']
[' Loewe', ' Wagner', ' Debussy']
[' Field', ' Chopin', ' Debussy']
[' Reicha', ' Liszt', ' Debussy']
[' Adam', ' Massenet', ' Debussy']
[' Gade', ' Grieg', ' Debussy']
[' Indy', ' Satie', ' Debussy']
[' Duparc', ' Chausson', ' Debussy']
[' Hasse', ' Bach, JS', ' Debussy']
[' Torelli', ' Bach, JS', ' Debussy']
[' Carissimi', ' Couperin, F', ' Debussy']
[' Lortzing', ' Wagner', ' Debussy']
[' Cavalli', ' Rameau', ' Debussy']
[' Zelenka', ' Bach, JS', ' Debussy']
[' Froberger', ' Bach, JS', ' Debussy']
[' Halevy', ' Wagner', ' Debussy']
[' Gretry', ' Franck', ' Debussy']
[' Czerny', ' Liszt', ' Debussy']
[' Raff', ' Tchaikovsky', ' Debussy']
[' Dargomizhsky', ' Rimsky-Korsakov', ' Debussy']
[' Reinecke', ' Grieg', ' Debussy']
[' Spontini', ' Wagner', ' Debussy']
[' Dussek', ' Chopin', ' Debussy']
[' Couperin, L', ' Couperin, F', ' Debussy']
[' Lalande', ' Couperin, F', ' Debussy']
[' Lotti', ' Bach, JS', ' Debussy']
[' Auber', ' Wagner', ' Debussy']
[' Bononcini', ' Rameau', ' Debussy']
[' Allegri', ' Liszt', ' Debussy']
[' Bohm', ' Bach, JS', ' Debussy']
[' Mehul', ' Wagner', ' Debussy']
[' Legrenzi', ' Bach, JS', ' Debussy']
[' Bruhns', ' Bach, JS', ' Debussy']
[' Hartmann, JPE', ' Grieg', ' Debussy']

3 steps
[' Haydn, J', ' Mozart, WA', ' Wagner', ' Debussy']
[' Mahler', ' Schoenberg', ' Ravel', ' Debussy']
[' Dvorak', ' Schoenberg', ' Ravel', ' Debussy']
[' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Purcell', ' Handel', ' Bach, JS', ' Debussy']
[' Bruckner', ' Strauss, R', ' Ravel', ' Debussy']
[' Strauss, J_Jr', ' Strauss, R', ' Ravel', ' Debussy']
[' Monteverdi', ' Schutz', ' Bach, JS', ' Debussy']
[' Scriabin', ' Stravinsky', ' Ravel', ' Debussy']
[' Bach, CPE', ' Mozart, WA', ' Wagner', ' Debussy']
[' Wolf', ' Schoenberg', ' Ravel', ' Debussy']
[' Byrd', ' Sweelinck', ' Bach, JS', ' Debussy']
[' Berg', ' Gershwin', ' Ravel', ' Debussy']
[' Webern', ' Stravinsky', ' Ravel', ' Debussy']
[' Offenbach', ' Bizet', ' Tchaikovsky', ' Debussy']
[' Lasso', ' Charpentier, M-A', ' Couperin, F', ' Debussy']
[' Josquin', ' Palestrina', ' Bach, JS', ' Debussy']
[' Milhaud', ' Gershwin', ' Ravel', ' Debussy']
[' Reger', ' Schoenberg', ' Ravel', ' Debussy']
[' Mascagni', ' Puccini', ' Ravel', ' Debussy']
[' Scarlatti, A', ' Handel', ' Bach, JS', ' Debussy']
[' Dowland', ' Sweelinck', ' Bach, JS', ' Debussy']
[' Busoni', ' Schoenberg', ' Ravel', ' Debussy']
[' Lehar', ' Puccini', ' Ravel', ' Debussy']

[' Boccherini', ' Mozart, WA', ' Wagner', ' Debussy']
[' Bach, JC', ' Mozart, WA', ' Wagner', ' Debussy']
[' Glazunov', ' Stravinsky', ' Ravel', ' Debussy']
[' Boito', ' Verdi', ' Mussorgsky', ' Debussy']
[' Pergolesi', ' Mozart, WA', ' Wagner', ' Debussy']
[' Gabrieli, G', ' Schutz', ' Bach, JS', ' Debussy']
[' Victoria', ' Palestrina', ' Bach, JS', ' Debussy']
[' Cowell', ' Gershwin', ' Ravel', ' Debussy']
[' Haydn, M', ' Mozart, WA', ' Wagner', ' Debussy']
[' Tartini', ' Mozart, WA', ' Wagner', ' Debussy']
[' Machaut', ' Stravinsky', ' Ravel', ' Debussy']
[' Ponchielli', ' Puccini', ' Ravel', ' Debussy']
[' Gesualdo', ' Stravinsky', ' Ravel', ' Debussy']
[' Zemlinsky', ' Schoenberg', ' Ravel', ' Debussy']
[' Cimarosa', ' Mozart, WA', ' Wagner', ' Debussy']
[' Praetorius, M', ' Schutz', ' Bach, JS', ' Debussy']
[' Scheidt', ' Buxtehude', ' Bach, JS', ' Debussy']
[' Paisiello', ' Mozart, WA', ' Wagner', ' Debussy']
[' Giuliani', ' Schubert', ' Liszt', ' Debussy']
[' Caldara', ' Handel', ' Bach, JS', ' Debussy']
[' Gabrieli, A', ' Pachelbel', ' Bach, JS', ' Debussy']
[' Galuppi', ' Salieri', ' Liszt', ' Debussy']
[' Locatelli', ' Paganini', ' Chopin', ' Debussy']
[' Fux', ' Mozart, WA', ' Wagner', ' Debussy']
[' Marenzio', ' Schutz', ' Bach, JS', ' Debussy']
[' Kuhlau', ' Hartmann, JPE', ' Grieg', ' Debussy']
[' Geminiani', ' Paganini', ' Chopin', ' Debussy']
[' Quantz', ' Hasse', ' Bach, JS', ' Debussy']
[' Mercadante', ' Verdi', ' Mussorgsky', ' Debussy']
[' Schein', ' Schutz', ' Bach, JS', ' Debussy']
[' Bull', ' Sweelinck', ' Bach, JS', ' Debussy']
[' Hassler', ' Schutz', ' Bach, JS', ' Debussy']
[' Willaert', ' Palestrina', ' Bach, JS', ' Debussy']
[' Stradella', ' Handel', ' Bach, JS', ' Debussy']
[' Morales', ' Palestrina', ' Bach, JS', ' Debussy']
[' Cabezon', ' Sweelinck', ' Bach, JS', ' Debussy']
[' Caccini', ' Frescobaldi', ' Bach, JS', ' Debussy']
[' Schmitt', ' Stravinsky', ' Ravel', ' Debussy']
[' Viotti', ' Mozart, WA', ' Wagner', ' Debussy']
[' Rore', ' Palestrina', ' Bach, JS', ' Debussy']
[' Mayr', ' Rossini', ' Wagner', ' Debussy']
[' Sammartini, GB', ' Mozart, WA', ' Wagner', ' Debussy']
[' Danzi', ' Weber', ' Wagner', ' Debussy']
[' Muffat', ' Handel', ' Bach, JS', ' Debussy']
[' Jommelli', ' Mozart, WA', ' Wagner', ' Debussy']
[' Boieldieu', ' Weber', ' Wagner', ' Debussy']
[' Philips', ' Sweelinck', ' Bach, JS', ' Debussy']
[' Piccinni', ' Mozart, WA', ' Wagner', ' Debussy']
[' Stamitz, J', ' Mozart, WA', ' Wagner', ' Debussy']
[' Martini', ' Mozart, WA', ' Wagner', ' Debussy']

4 steps

[' Bartok', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Britten', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Prokofiev', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Ives', ' Cowell', ' Gershwin', ' Ravel', ' Debussy']
[' Hindemith', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Villa-Lobos', ' Milhaud', ' Gershwin', ' Ravel', ' Debussy']
[' Weill', ' Zemlinsky', ' Schoenberg', ' Ravel', ' Debussy']
[' Smetana', ' Dvorak', ' Schoenberg', ' Ravel', ' Debussy']
[' Nielsen', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Delius', ' Schmitt', ' Stravinsky', ' Ravel', ' Debussy']
[' Dufay', ' Josquin', ' Palestrina', ' Bach, JS', ' Debussy']
[' Gibbons, O', ' Purcell', ' Handel', ' Bach, JS', ' Debussy']
[' Krenek', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Herbert', ' Dvorak', ' Schoenberg', ' Ravel', ' Debussy']
[' Tallis', ' Purcell', ' Handel', ' Bach, JS', ' Debussy']
[' Strauss, J_Sr', ' Strauss, J_Jr', ' Strauss, R', ' Ravel', ' Debussy']
[' Strauss, Jos', ' Strauss, J_Jr', ' Strauss, R', ' Ravel', ' Debussy']
[' Ockeghem', ' Josquin', ' Palestrina', ' Bach, JS', ' Debussy']

[' Biber', ' Muffat', ' Handel', ' Bach, JS', ' Debussy']
[' Morley', ' Dowland', ' Sweelinck', ' Bach, JS', ' Debussy']
[' Blow', ' Purcell', ' Handel', ' Bach, JS', ' Debussy']
[' Isaac', ' Webern', ' Stravinsky', ' Ravel', ' Debussy']
[' Koechlin', ' Milhaud', ' Gershwin', ' Ravel', ' Debussy']
[' Tomkins', ' Philips', ' Sweelinck', ' Bach, JS', ' Debussy']
[' Flotow', ' Offenbach', ' Bizet', ' Tchaikovsky', ' Debussy']
[' Marcello', ' Martini', ' Mozart, WA', ' Wagner', ' Debussy']
[' Dittersdorf', ' Haydn, J', ' Mozart, WA', ' Wagner', ' Debussy']
[' Lyadov', ' Scriabin', ' Stravinsky', ' Ravel', ' Debussy']
[' Arensky', ' Scriabin', ' Stravinsky', ' Ravel', ' Debussy']
[' Locke', ' Purcell', ' Handel', ' Bach, JS', ' Debussy']
[' Schreker', ' Berg', ' Gershwin', ' Ravel', ' Debussy']
[' Lawes', ' Purcell', ' Handel', ' Bach, JS', ' Debussy']
[' Taneyev', ' Scriabin', ' Stravinsky', ' Ravel', ' Debussy']
[' Obrecht', ' Josquin', ' Palestrina', ' Bach, JS', ' Debussy']
[' Holborne', ' Dowland', ' Sweelinck', ' Bach, JS', ' Debussy']
[' Taverner', ' Byrd', ' Sweelinck', ' Bach, JS', ' Debussy']
[' Billings', ' Cowell', ' Gershwin', ' Ravel', ' Debussy']
[' Gombert', ' Morales', ' Palestrina', ' Bach, JS', ' Debussy']
[' Schmelzer', ' Fux', ' Mozart, WA', ' Wagner', ' Debussy']
[' Ruggles', ' Cowell', ' Gershwin', ' Ravel', ' Debussy']
[' Clemens', ' Byrd', ' Sweelinck', ' Bach, JS', ' Debussy']

5 steps

[' Rachmaninov', ' Prokofiev', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Copland', ' Britten', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Sullivan', ' Herbert', ' Dvorak', ' Schoenberg', ' Ravel', ' Debussy']
[' Kodaly', ' Bartok', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Walton', ' Britten', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Holst', ' Britten', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Varese', ' Ruggles', ' Cowell', ' Gershwin', ' Ravel', ' Debussy']
[' Humperdinck', ' Weill', ' Zemlinsky', ' Schoenberg', ' Ravel', ' Debussy']
[' Bridge', ' Britten', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Gliere', ' Prokofiev', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Ireland', ' Britten', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Berwald', ' Nielsen', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Weelkes', ' Tomkins', ' Philips', ' Sweelinck', ' Bach, JS', ' Debussy']
[' Miaskovsky', ' Prokofiev', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Binchois', ' Dufay', ' Josquin', ' Palestrina', ' Bach, JS', ' Debussy']
[' Svendsen', ' Nielsen', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Chadwick', ' Ives', ' Cowell', ' Gershwin', ' Ravel', ' Debussy']
[' Sinding', ' Delius', ' Schmitt', ' Stravinsky', ' Ravel', ' Debussy']
[' Dunstable', ' Dufay', ' Josquin', ' Palestrina', ' Bach, JS', ' Debussy']

6 steps

[' Sibelius', ' Walton', ' Britten', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Vaughan Williams', ' Holst', ' Britten', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Elgar', ' Walton', ' Britten', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Stanford', ' Holst', ' Britten', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Widor', ' Varese', ' Ruggles', ' Cowell', ' Gershwin', ' Ravel', ' Debussy']
[' Thomson', ' Copland', ' Britten', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Chavez', ' Copland', ' Britten', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Rheinberger', ' Chadwick', ' Ives', ' Cowell', ' Gershwin', ' Ravel', ' Debussy']
[' Medtner', ' Rachmaninov', ' Prokofiev', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']

7 steps

[' Ponce', ' Chavez', ' Copland', ' Britten', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Parry', ' Vaughan Williams', ' Holst', ' Britten', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Butterworth', ' Vaughan Williams', ' Holst', ' Britten', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']
[' Goldmark', ' Sibelius', ' Walton', ' Britten', ' Shostakovich', ' Gershwin', ' Ravel', ' Debussy']

Appendix 2: The CSC similarity measure computed as the Pearson correlation coefficient *r* between Boolean vectors

From discussion in Section 3, the centralised cosine similarity measure is:

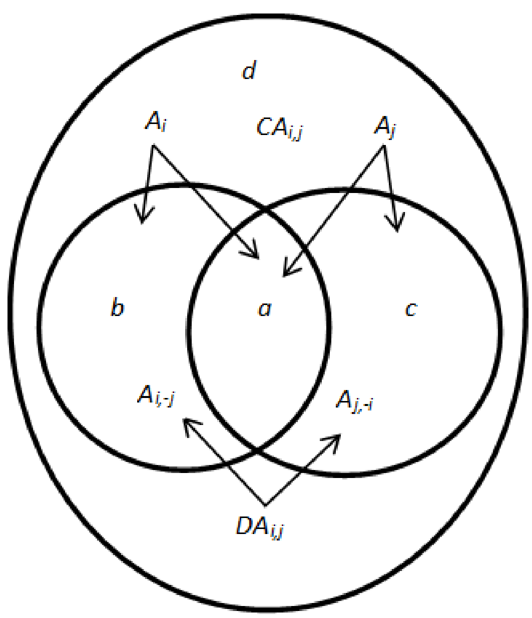
(A2.1) $$CSC_{i,j} = (ad - bc) \Big/ \sqrt{(a+b)(c+d)(a+c)(b+d)}$$

where *a, b, c, d* are counts of attributes for the pair of composers (*i,j*) as illustrated in the count table below. Attributes are the 298 ecological characteristics, plus 42 style influences and 500 personal musical influences. So the total number of attributes is $n_k = 840$. We can also represent the issue with a set diagram as in Figure A2.1. Note that if we focused only on some attributes, say, the 298 ecological features, then $n_k = 298$.

Table A2.1: 2 by 2 frequency table for Presence/Absence of composer attributes using counts

		Composer *j*		
Composer *i*		Presence	Absence	Total
	Presence	*a*	*b*	*a*+*b*
	Absence	*c*	*d*	*c*+*d*
	Total	*a*+*c*	*b*+*d*	n_k

Figure A2.1: Set diagram (see text for explanation)



Although the centralised cosine formula in Eq. (A2.1) is based on counts of presence/absence of attributes as Table A2.1 and Figure A2.1 underlie, we can obtain this formula, perhaps more conventionally, as follows. Consider each composer *i* as a $n_k \times 1$ Boolean vector of 0's and 1's. That is, if an attribute *k* among the n_k possible attributes belongs to *i*, then the k^{th} component of the vector corresponding to composer *i* is set equal to 1, otherwise it is set equal to 0. The centralised cosine similarity measure for a pair of composers (i, j), each represented by their own Boolean vectors B_i and B_j, can then be computed as:

(A2.2)
$$CSC_{i,j} = \frac{\sum_{k=1}^{n_k}\left(B_{k,i} - \overline{B_i}\right)\times\left(B_{k,j} - \overline{B_j}\right)}{\sqrt{\sum_{k=1}^{n_k}\left(B_{k,i} - \overline{B_i}\right)^2}\sqrt{\sum_{k=1}^{n_k}\left(B_{k,j} - \overline{B_j}\right)^2}},$$

where subscript k in $B_{k,i}$ indicates the k^{th} component (of value 1 or 0) of vector B_i. Assuming that $\overline{B_i} = (1/n_k)\sum_{k=1}^{n_k} B_{k,i}$ and $\overline{B_j} = (1/n_k)\sum_{k=1}^{n_k} B_{k,j}$, the centralised cosine measure is the cosine measure computed on the centralised vectors, with respect to the mean (average) vectors. That is, without centralisation, we would have the ordinary cosine measure:

(A2.3)
$$COS_{i,j} = \frac{\sum_{k=1}^{n_k} B_{k,i} \times B_{k,j}}{\sqrt{\sum_{k=1}^{n_k}\left(B_{k,i}\right)^2}\sqrt{\sum_{k=1}^{n_k}\left(B_{k,j}\right)^2}}.$$

With centralisation, Eq. (A2.2) is, in fact, the formula for the Pearson correlation coefficient applied, here, on Boolean vectors. We can establish the link between (A2.2) and (A2.1) as follows. Note that in terms of our notation in the presence-absence Table A2.1 (and given that Boolean vectors are made of 0's and 1's) we have that:

$$\sum_{k=1}^{n_k} B_{k,i} = a + b,$$
$$\sum_{k=1}^{n_k} B_{k,j} = a + c,$$
$$\sum_{k=1}^{n_k} B_{k,i} \times B_{k,j} = a,$$
$$\sum_{k=1}^{n_k}\left(B_{k,i}\right)^2 = a + b,$$
$$\sum_{k=1}^{n_k}\left(B_{k,j}\right)^2 = a + c,$$
$$\overline{B_i} = \frac{a+b}{n_k},$$
$$\overline{B_j} = \frac{a+c}{n_k}.$$

Rewriting Eq. (A2.2) as:

$$CSC_{i,j} = \frac{\sum_{k=1}^{n_k}\left(B_{k,i}B_{k,j} - B_{k,i}\overline{B_j} - \overline{B_i}B_{k,j} + \overline{B_i}\,\overline{B_j}\right)}{\sqrt{\sum_{k=1}^{n_k}\left(B_{k,i}^2 - 2B_{k,i}\overline{B_i} + \overline{B_i}^2\right)}\sqrt{\sum_{k=1}^{n_k}\left(B_{k,j}^2 - 2B_{k,j}\overline{B_j} + \overline{B_j}^2\right)}},$$

and substituting the notations above, we obtain after some algebraic simplifications, that:
(A2.4)

$$CSC_{i,j} = \frac{\sum_{k=1}^{n_k}\left(B_{k,i} - \overline{B_i}\right)\times\left(B_{k,j} - \overline{B_j}\right)}{\sqrt{\sum_{k=1}^{n_k}\left(B_{k,i} - \overline{B_i}\right)^2}\sqrt{\sum_{k=1}^{n_k}\left(B_{k,j} - \overline{B_j}\right)^2}} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}.$$

Doing the same type of algebraic simplifications for the cosine measure in Eq. (A2.3), we obtain:

(A2.5)
$$COS_{i,j} = \frac{a}{\sqrt{(a+b)(a+c)}}.$$

The ordinary (non-centralised) cosine similarity measure (also known as the Salton's measure) is a statistic familiar to bibliometrics and scientometrics. The idea was mathematically formalized by Sen and Gan (1983) and later extended by Glänzel and Czerwon (1996) who also applied the methodology. Egghe and Leydesdorff (2009) show that there is no pure functional relation between the Pearson correlation coefficient *r* (or *CSC*), and the ordinary cosine measure (*COS*). However, they establish that the cloud of points (*COS, r*) can be described by a sheaf of increasing straight lines whose slope decreases, the higher the straight line is in the sheaf. Smith et al. (2015) show that there is a quadratic relation between the centralised cosine measure and the chi-square statistic/binomial index of dispersion (*BID*) (a traditional association measure) so that:

(A2.6) $$BID_{i,j} = nCSC_{i,j}{}^{2} = n(ad-bc)^2 \big/ \left[(a+b)(c+d)(a+c)(b+d)\right].$$

See Hayek (1994) for formulas of up to 46 coefficients on measures of association, including similarity coefficients (Simpson, Kulczynski, Dice, Jaccard, etc.), matching coefficients (Sokal-Sneath, Russell-Rao, Sneath, etc.) and traditional association measures (chi-square statistic/binomial index of dispersion, coefficient of mean square contingency, Phi coefficient, Pearce, etc.) Smith and Georges (2014) compare some of these indices for the *CMN* database.

Some intuition for the centralised cosine formula in Eq. (A2.4) may be provided as follows using Table A2.1. If no association exists between composers *i* and *j*, the proportion of attributes of composer *i* in the overall database of attributes (n_k) should be equal to 1) the proportion of attributes of *i* which are also attributes of *j* in the overall set of attributes of *j* and 2) the proportion of attributes of *i* which are not attributes of *j* in the overall set of attributes which do not belong to *j*. That is,

(A2.6) $$(a+b)/n_k = a/(a+c) = b/(b+d).$$

Assuming for example that *i* = J.S. Bach and *j* = Mozart, we say that there is no association between Bach and Mozart in the case where, say, 5% of the total attributes (n_k) in the database relate to Bach (a+b) (the first term) and then, when observing attributes of Mozart (a+c), we find that 5% of these attributes are also attributes of Bach (a) (the second term), and that 5% of those attributes that do not relate to Mozart (b+d) nevertheless relate to Bach (b) (the third term). However, a positive association between Bach and Mozart is inferred if we find that the first, second and third terms have values of, say, 5%, 9%, and 1%.

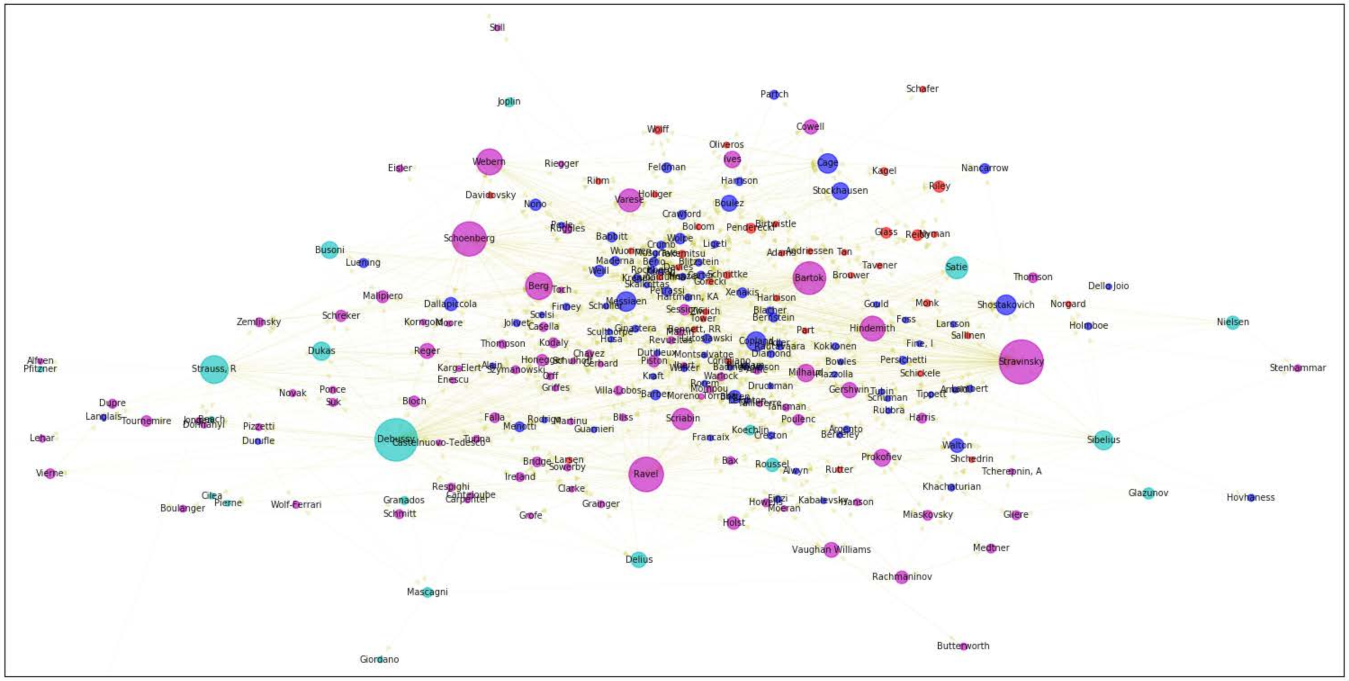
We can rewrite Eq. (A2.6) as: $(a+b)/n_k = a/(a+c) = b/(b+d)$, so that $a = (a+b)(a+c)/n_k$ or: $a/n_k = ((a+b)/n_k)((a+c)/n_k)$; that is, when two composers are independent (lack of association), the proportion or frequency of joint attributes (a/n_k) is equivalent to the product of the proportions $(a+b)/n_k$ and $(a+c)/n_k$ (that is, the proportion of attributes in the database which relate to *i* and the proportion of attributes that relate to *j*). If the observed frequency is greater than the one expected under independence, then the two composers may be said to be positively associated. Thus, if composers *i* and *j* are associated, then: $a \neq (a+b)(a+c)/n_k$, and the difference could be written as:

(A2.7) $$D = a-(a+b)(a+c)/n_k = (a/n_k)(n_k-a-b-c)-bc/n_k = (ad-bc)/n_k.$$

This term *D*, or some variation of it, is found in the formula for calculating the centralised cosine as in Eq. (A2.4) or the chi-square statistic/binomial index of dispersion as in Eq. (A2.6).

Graphical Appendix – Supplementary graphs

Figure A1: 20th Century composers' network. Colour code by age groups. ForceAtlas2 algorithm



Note: Cyan for the late Romantic composers; magenta for those born 1870-1899; blue for those born 1900-1929; and red for those born in 1930 and after. See Footnote 9 for further explanations.

Figure A2: MDS in a '3-dimension' graph

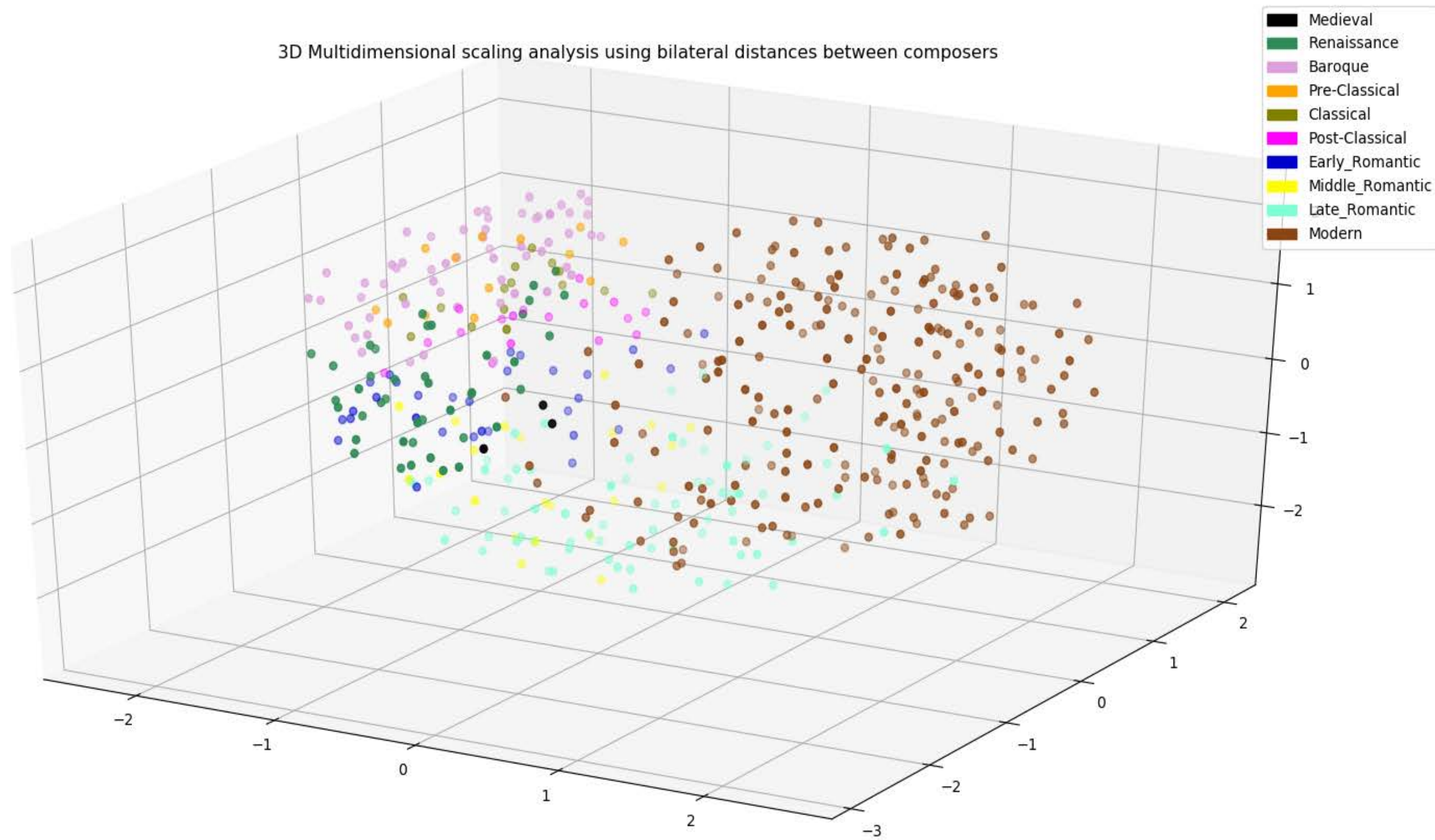


Figure A3: MDS using bilateral distances between 20th century composers

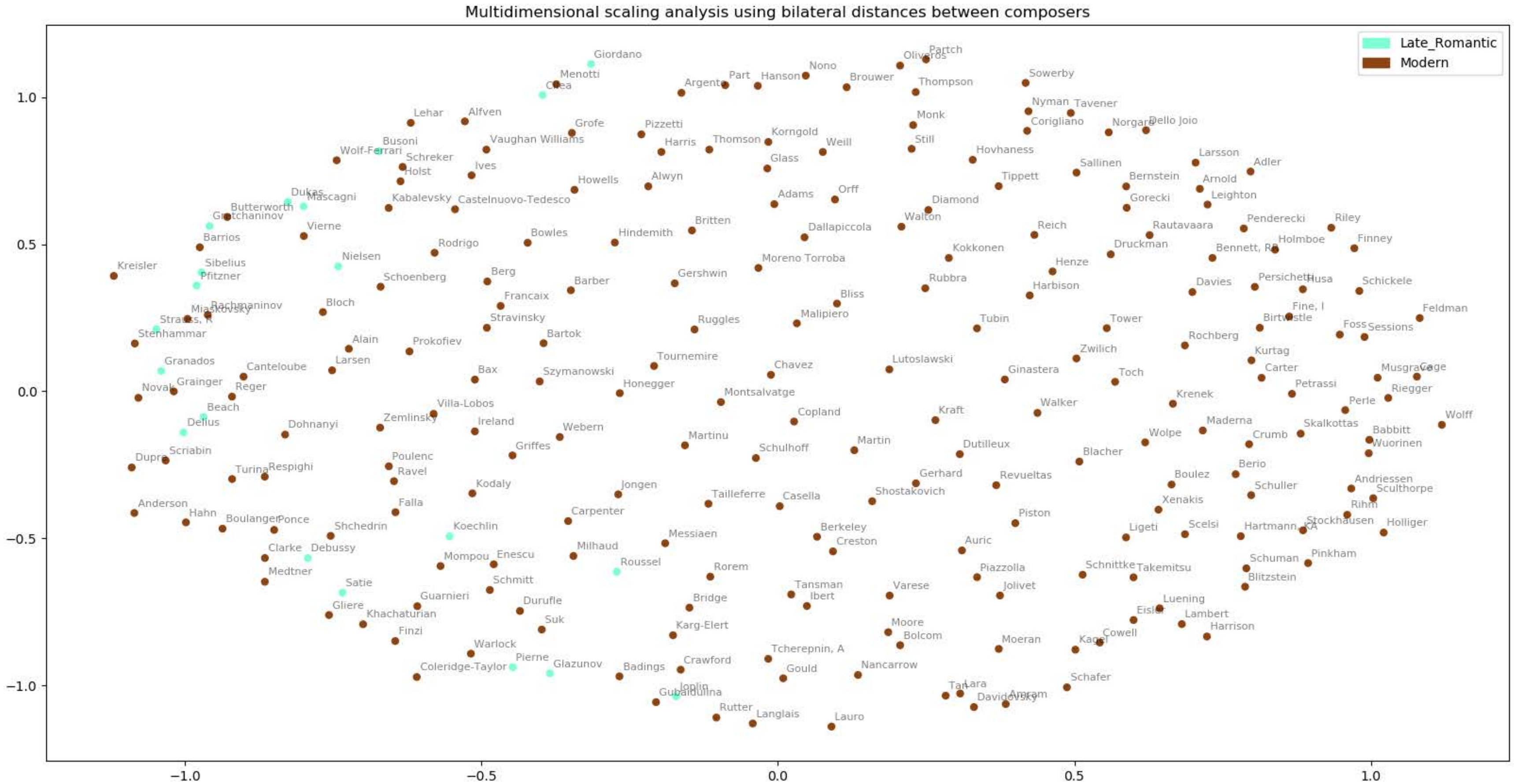


Figure A4: SVM on MDS map for 20th century composers and overfitting. The ‘microcosms’ of 20th century classical music

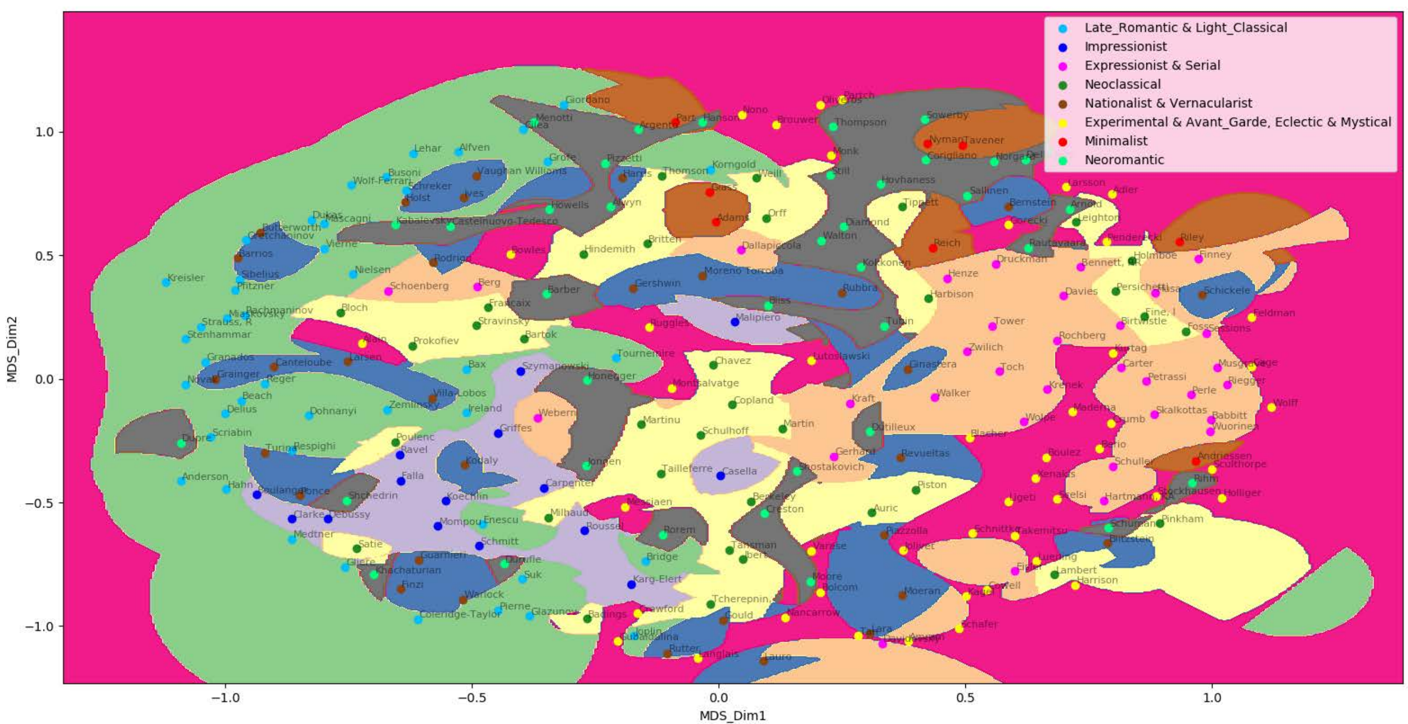
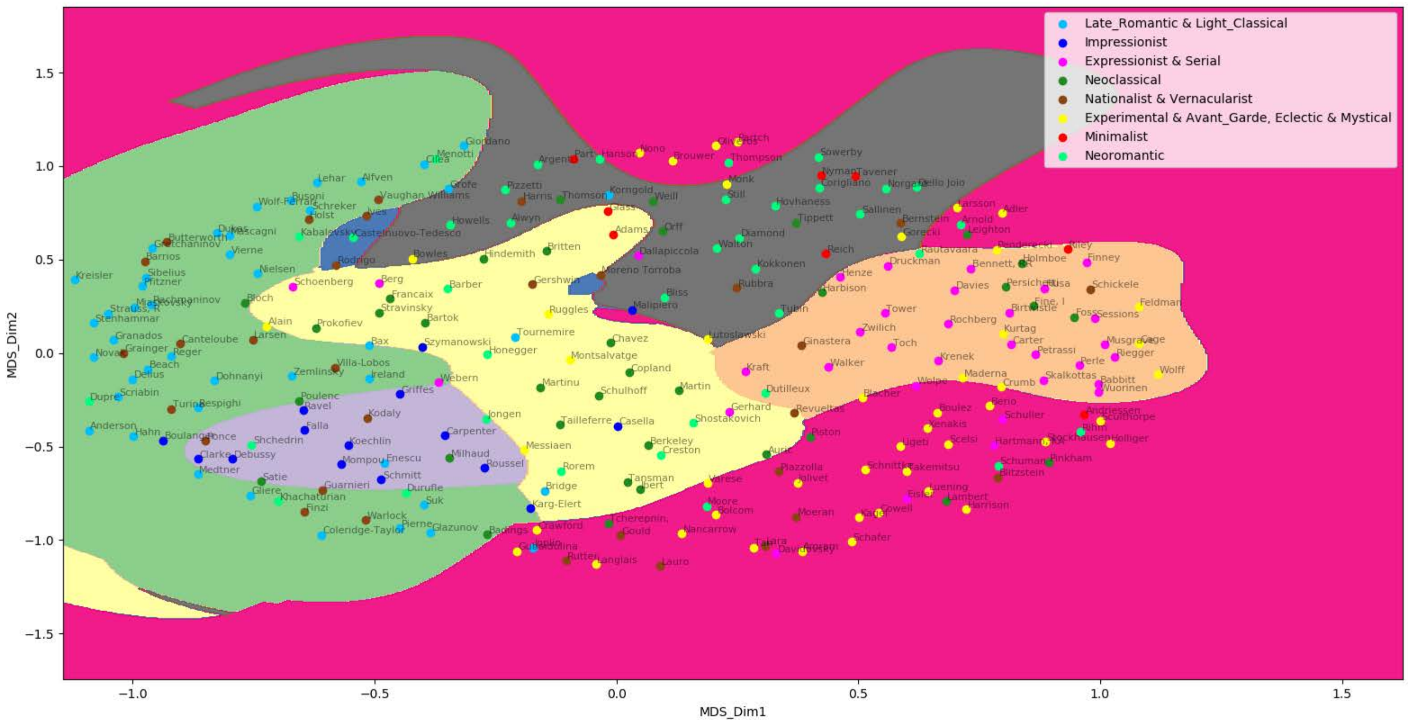
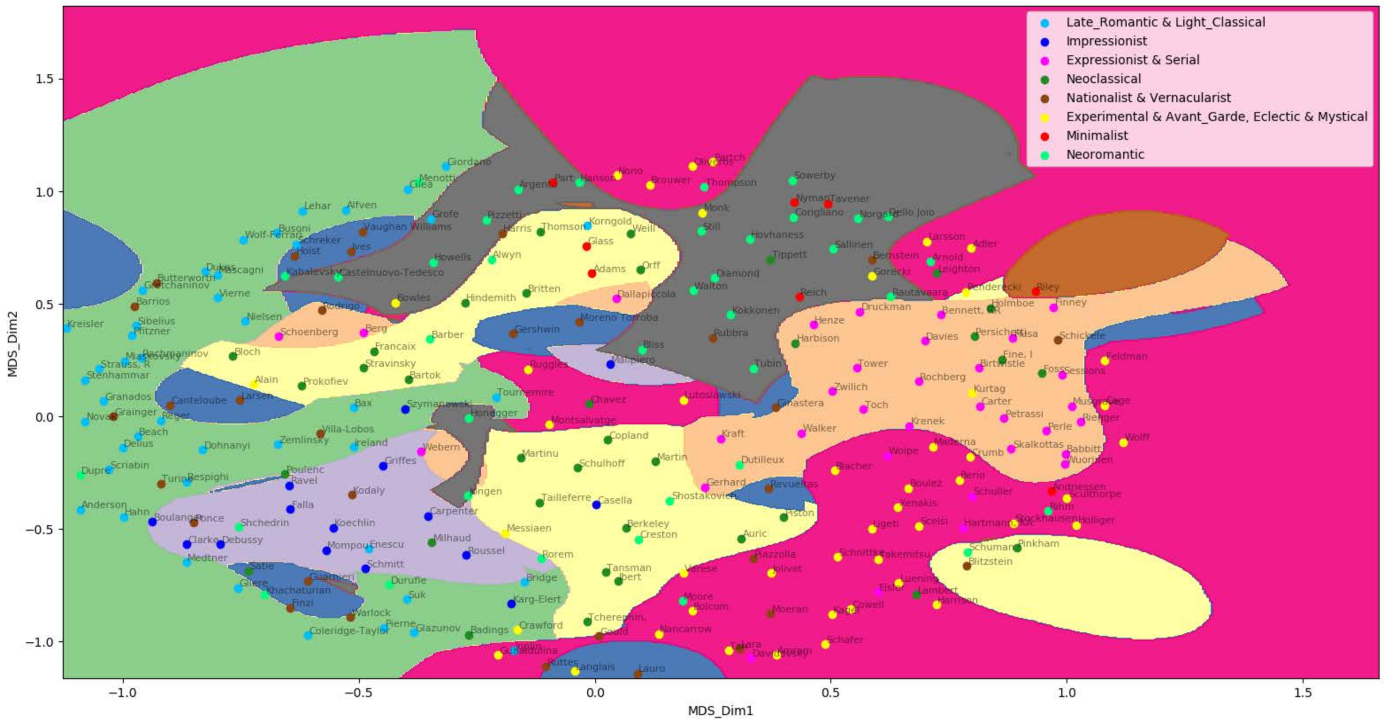


Figure A5: SVM on MDS map for 20th century composers and less overfitting



Note: The graph misses minimalists and nationalists

Figure A6: SVM on MDS map for 20th century composers – The ‘in between’ case



Note: The graphs shows zones with minimalists and nationalists

Figure A7: Clusters of 20th century composers (based on Figure 11 in text)

	Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5
13	Granados	114	Chavez	163	Cage	149	Langlais	214	Shchedrin
0	Debussy	115	Revueltas	199	Feldman	162	Alain	131	Khachaturian
38	Ravel	109	Lara	224	Wolff	24	Tournemire	134	Kabalevsky
36	Hahn	64	Ponce	205	Musgrave	32	Jongen	117	Tcherepnin_A
3	Pierne	150	Guarnieri	74	Riegger	91	Boulanger	58	Miaskovsky
9	Dukas	237	Brouwer	174	Perle	129	Durufle	82	Prokofiev
5	Gretchaninov	72	Barrios	223	Davidovsky	21	Vierne	146	Shostakovich
8	Glazunov	179	Lauro	176	Babbitt	75	Dupre	144	Tubin
28	Scriabin	2	Mascagni	235	Wuorinen	132	Berkeley	156	Holmboe
30	Rachmaninov	12	Cilea	125	Crawford	164	Francaix	155	Larsson
40	Gliere	15	Giordano	175	Ginastera	79	Martinu	228	Sallinen
34	Ives	201	Argento	105	Gerhard	80	Martin	158	Schuman
7	Nielsen	43	Wolf-Ferrari	153	Carter	177	Dutilleux	116	Thompson
6	Sibelius	161	Menotti	190	Ligeti	108	Tansman	103	Hanson
26	Stenhammar	124	Finzi	221	Davies	118	Auric	112	Harris
4	Strauss_R	35	Holst	222	Birtwistle	81	Ibert	209	Leighton
19	Pfitzner	61	Grainger	140	Scelsi	89	Tailleferre	126	Rubbra
31	Reger	27	Vaughan_Williams	197	Schuller	191	Rorem	186	Kokkonen
11	Busoni	94	Warlock	121	Krenek	17	Koechlin	96	Moeran
54	Medtner	73	Butterworth	130	Wolpe	90	Mompou	215	Norgard
16	Beach	76	Clarke	135	Dallapiccola	85	Milhaud	127	Partch
46	Dohnanyi	1	Delius	198	Henze	113	Poulenc	178	Harrison
25	Zemlinsky	52	Canteloube	220	Schnittke	10	Satie	106	Cowell
48	Schreker	70	Griffes	143	Hartmann_KA	142	Lambert	160	Hovhaness
71	Berg	165	Montsalvatge	78	Toch	59	Stravinsky	238	Andriessen
33	Schoenberg	51	Ireland	181	Rochberg	86	Honegger	245	Adams
66	Webern	62	Szymanowski	207	Crumb	67	Varese	227	Riley
20	Lehar	77	Villa-Lobos	202	Stockhausen	152	Messiaen	229	Reich
107	Korngold	99	Castelnuovo-Tedesco	187	Xenakis			231	Glass
47	Karg-Elert	84	Moreno_Torroba	196	Boulez			212	Gubaidulina
57	Enescu	42	Falla	208	Sculthorpe			243	Nyman
23	Schmitt	63	Turina	139	Jolivet			248	Tan
50	Bridge	123	Rodrigo	210	Takemitsu			102	Thomson
41	Coleridge-Taylor	49	Respighi	182	Maderna			241	Monk
22	Novak	55	Pizzetti	194	Nono			216	Oliveros
29	Alfven	60	Kodaly	195	Berio			219	Schafer
37	Suk	98	Orff	213	Kagel			101	Sowerby
68	Bax	97	Hindemith	240	Holliger			92	Moore

95	Schulhoff

53	Bloch
56	Bartok
157	Barber
167	Britten
138	Tippett
128	Walton
168	Lutoslawski

247	Rihm
225	Part
217	Penderecki
200	Kurtag
218	Gorecki

169	Dello_Joio
232	Bolcom
14	Joplin
100	Still
246	Larsen
171	Fine_I
233	Harbison
45	Ruggles
193	Kraft
166	Nancarrow
183	Piazzolla
120	Weill
170	Gould
180	Bernstein
119	Copland
159	Bowles
110	Gershwin
211	Amram
111	Eisler
141	Blitzstein
44	Carpenter
88	Grofe
83	Bliss
184	Arnold
147	Creston
87	Howells
145	Alwyn
203	Rautavaara
236	Tower
239	Zwilich
206	Druckman
230	Bennett_RR
173	Diamond
234	Corigliano
204	Adler
242	Tavener
192	Pinkham
244	Rutter
39	Kreisler
148	Finney
154	Anderson

226	Schickele
172	Persichetti
188	Foss
151	Badings
185	Husa
122	Luening
65	Malipiero
69	Casella
137	Petrassi
136	Skalkottas
189	Walker
93	Piston
104	Sessions

Acknowledgements

I thank Charles H. Smith for his comments on previous drafts, his numerous advices and suggestions, and for giving me access to data of the *CMN* website.

References

Buja, A., D. Swayne, M. Littman, N. Dean, H. Hofmann, and L. Chen (2008). Data visualisation with Multidimensional Scaling. *Journal of Computational and Graphical Statistics*, *17*(2): 444-474.

Chippada, B. (2017). ForceAtlas2 for Python. https://github.com/bhargavchippada/forceatlas2. (Accessed August 18, 2020.)

Crossley, N., S. McAndrew, and P. Widdop. (2015). *Social Networks and Music Worlds*. Routledge, NewYork.

Egghe and Leydesdorff (2009). The relation between Pearson's correlation coefficient *r* and Salton's cosine measure. *Journal of the American Society for Information Science & Technology, 60*(5): 1027-1036.

Georges, P. (2017). Western classical music development: A statistical analysis of composers similarity, differentiation and evolution. *Scientometrics*, *112*(1): 21-53.

Georges, P. and N. Nguyen. (2019). Visualizing Music Similarity: Clustering and Mapping 500 Classical Music Composers. *Scientometrics*, *120*(3): 975-1003.

Giller, G. L. (2012). The statistical properties of random bitstreams and the sampling distribution of cosine similarity. https://www.researchgate.net/publication/253236957_The_Statistical_Properties_of_Random_Bitstreams_and_the_Sampling_Distribution_of_Cosine_Similarity. (Accessed 30 July 2020.)

Glänzel, W. and H. J. Czerwon. (1996). A new methodological approach to bibliographic coupling and its application to the national, regional, and institutional level. *Scientometrics, 37*(2), 195–221.

Griffiths, P. (1978). *A Concise History of Modern Music: From Debussy to Boulez*. Thames & Hudson, London.

Hayek, L.-A. C. (1994). Analysis of amphibian biodiversity data. In W. R. Heyer, M. A. Donnelly, R. W. McDiarmid, L.-A. C. Hayek, & M. S. Foster (Eds.), *Measuring and monitoring biological diversity: Standard methods for amphibians* (pp. 207–269). Washington, DC: Smithsonian Books.

Jackson, M. (2011). An overview of social networks and economic applications. In Handbook of Social Economics, Vol. 1, Chap. 12, 511-585.

Jacomy, M., T. Venturini, S. Heymann and M. Bastian. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualisation designed for the Gephi software. *PLoS ONE*, 9(6): e98679 June 10, 2014

Knees, P. and M. Schedl. (2016). *Music Similarity and Retrieval. An Introduction to Audio- and Web-based Strategies*. Springer-Verlag Berlin Heidelberg.

Khulusi, R., J. Kusnick, C. Meinecke, C. Gillmann, J. Focht, J. and S. Jänicke . (2020). A Survey on Visualizations for Musical Data. *Computer Graphics Forum.* Early and Open Access: 05 March 2020: doi:10.1111/cgf.13905

Ladd, J.R., J. Otis, C.N. Warren, and S. Weingart. (2017). Exploring and analyzing network data with Python. *The Programming Historian* (6) https://doi.org/10.46430/phen0064.

Magnuson, P. (2008). Sound Patterns. A Structural Examination of Tonality, Vocabulary, Texture, Sonorities and Time Organisation in Western Art Music. Academic.udayton.edu. (Accessed June 8, 2020).

McAndrew, S. and M. Everett. (2015). Music as Collective Invention: A Social Network Analysis of Composers. *Cultural Sociology* 9(*1*), 56-80.

Müller, M. (2015). *Fundamentals of Music Processing. Audio, Analysis, Algorithms, Applications*. Springer International Publishing Switzerland.

Orio, N. (2006). Music retrieval: A tutorial and review. *Foundations and Trends in Information Retrieval*, *1*(1), 1-90.

Pauls, H. (2014). *Two Centuries in One. Musical Romanticism and the Twentieth Century*. Ph.D. thesis, Rostock University of Music and Theatre.

Rutherford-Johnson, T. (2017). *Music after the Fall. Modern Composition and Culture since 1989.* University of California Press, Oakland, California.

Schedl, M., E. Gomez, and J. Urbano. (2014). Music Information Retrieval: Recent Developments and Applications. *Foundations and Trends in Information Retrieval*, *8*(2-3), 127-261.

Sen, S. K. and S. K. Gan. (1983). A mathematical extension of the idea of bibliographic coupling and its applications. *Annals of Library and Information Studies, 30*(2), 78–823.

Smith, C. H. (2000). The classical music navigator. http://people.wku.edu/charles.smith/music/ (accessed September 07, 2019)

Smith, C.H. and P. Georges. (2014). Composer similarities through 'The Classical Music Navigator': Similarity inference from composer influences. *Empirical Studies of the Arts*, *32(2)*: 205-29.

Smith, C.H. and P. Georges. (2015). Similarity indices for 500 classical music composers: Inferences from personal musical influences and 'ecological' measures. *Empirical Studies of the Arts*, *33(1)*: 61-94.

Smith, C.H., P. Georges, and N. Nguyen. (2015). Statistical tests for 'related records' search results. *Scientometrics*, *105*(3): 1665-1677.

Weiss, C. (2017). *Computational methods for tonality-based style analysis of classical music audio recordings* (Doctoral dissertation). Ilmenau University of Technology. Ilmenau, Germany.

Weiss, C., M. Mauch, S. Dixon, M. Müller. (2018). Investigating style evolution of Western classical music: A computational approach. *Musicae Scientiae*, OnlineFirst, First published online: March 7, 2018. https://doi.org/10.1177/1029864918757595.