Interactive Visual Profiling of Musicians

Stefan Jänicke, Josef Focht, and Gerik Scheuermann, *Member, IEEE*

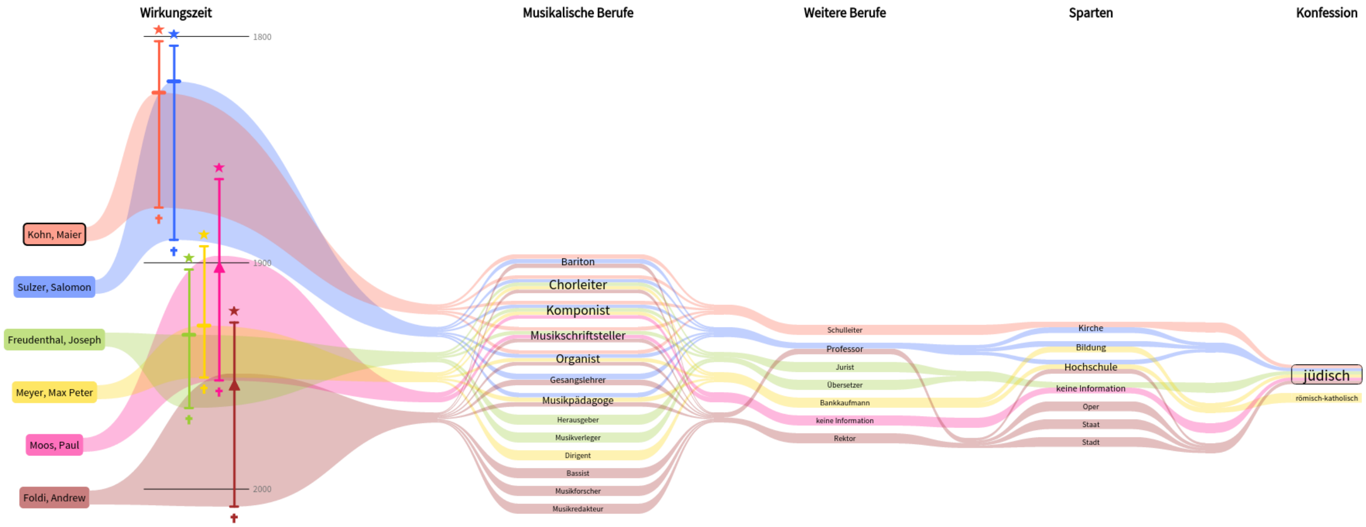

Fig. 1. The Column Explorer of our profiling system compares the activity time (Wirkungszeit) of musicians and shows correlations among various biographical information about musical professions (Musikalische Berufe), further professions (Weitere Berufe), divisions (Sparten) and denomination (Konfession). The profiling scenario for the Jewish cantor Maier Kohn using a mandatory *Jewish* (jüdisch) denomination detects other Jewish cantors as similar musicians with multifaceted interpretations of the cantor profession.

Abstract— Determining similar objects based upon the features of an object of interest is a common task for visual analytics systems. This process is called *profiling*, if the object of interest is a person with individual attributes. The profiling of musicians similar to a musician of interest with the aid of visual means became an interesting research question for musicologists working with the Bavarian Musicians Encyclopedia Online. This paper illustrates the development of a visual analytics profiling system that is used to address such research questions. Taking musicological knowledge into account, we outline various steps of our collaborative digital humanities project, priority (1) the definition of various measures to determine the similarity of musicians' attributes, and (2) the design of an interactive profiling system that supports musicologists in iteratively determining similar musicians. The utility of the profiling system is emphasized by various usage scenarios illustrating current research questions in musicology.

Index Terms—visual analytics, profiling system, musicians database visualization, digital humanities, musicology

1 Introduction

The digitization age changed strategies and methods to gain knowledge in the humanities essentially. In terms of philology, humanities approaches were traditionally oriented primarily on language and text. As retrieval strategies in printed encyclopedias were mostly based on the alphabetical order of contained names, the now available access to large relational databases provides the opportunity for humanities scholars to filter groups of entities not only based on names, but also on various other data facets. But the pure access to a database containing a multitude of information also reveals limitations when it comes to investigating concrete research questions beyond only showing lists of entries that match a given database query. Especially for humanities scholars who are not used to exploit the cardinality of query languages, it becomes hard to navigate large databases and filter results accordingly [25]. Also, with experiences of the so-called visualistic turn [45] in the early 1990's, humanities scholars more and more wished to explore complex issues not only based on texts, but also based on images.

An example of this process is given by the Bavarian Musicians Encyclopedia Online (German: "Bayerisches Musiker-Lexikon Online (BMLO)") [1], a powerful web-based interface for musicologists, as it provides access to information about around 30,000 musicians, extracted from various digitized printed media. Being an invaluable tool in the daily work of musicologists, this innovative access to music history heritage raised new research questions in musicology. Besides the visualization of a musician's profile and profile comparison, the major desire was the ability to find similar musicians based upon the attributes of a musician of interest, in short: *profiling*.

In order to develop a valuable visual analytics system capable of determining similar musicians interactively, we closely collaborated with those musicologists who originally raised the profiling research question. Initially, we discussed possible profiling tasks and various aspects of how the musicologists imagined profiling workflows. We transformed the provided data for profiling purposes, mapped required musicians' attributes to visual features, and defined similarity measures to be used as the basis for profiling similar musicians. In summary, our contributions to the visualization community are:

- *Stefan Jänicke and Gerik Scheuermann are with Image and Signal Processing Group, Institute for Computer Science, Leipzig University, Germany. E-mail: {stjaenicke,scheuermann}@informatik.uni-leipzig.de.*
- *Josef Focht is with Museum of Musical Instruments, Institute for Musicology, Leipzig University, Germany. E-mail: josef.focht@uni-leipzig.de.*

Manuscript received 31 Mar. 2015; accepted 1 Aug. 2015; date of publication 20 Aug. 2015; date of current version 25 Oct. 2015. For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org. Digital Object Identifier no. 10.1109/TVCG.2015.2467620

- **The similarity of person attributes**: For various attributes of musicians, we designed similarity measures in accordance to musicological conceptions.
- **Profiling system**: Based upon eight similarity measures and a semi-automatic weighting of attribute dimensions, we designed a visual analytics system that is used by musicologists to iteratively search for musicians similar to a musician of interest.
- **Column Explorer**: For the comparison of temporal and textual metadata of musicians, we provide a column based representation that borrows ideas from Jigsaw's list view and Parallel Tag Clouds in order to explore correlations among various attributes.
- **Temporal uncertainty**: We consider existent temporal uncertainties when fixing a musician's activity time and provide a design that communicates these uncertainties to the musicologists.

We emphasize the utility of our visual analytics profiling system for musicologists by providing various usage scenarios. In a storytelling style, each scenario exemplifies how the profiling system can be used to discover unexpected insights. Additionally, we report experiences gained during our digital humanities project. This includes the iterative evaluation of our profiling system with musicologists, limitations due to the nature of humanities data and future prospects.

2 Related Work

Our work on a profiling system for musicians is related to three different aspects of research. As the usual method of profiling is based upon similarities to a given subject, we first take a look at recommendation systems, where recommendations are determined in dependency on an object of interest. The second paragraph lists visual analytics and information visualization papers that (1) are motivated by similar research questions to ours without a digital humanities background, or (2) present similar visualization techniques to the ones our profiling system is composed of. Finally, we illustrate related visualizations motivated from research questions in digital humanities.

2.1 Recommendation Systems

Tapestry, the first recommender system developed in 1992, was based on collaborative filtering [26]. The system aggregates recommendations provided by users, and directs them to appropriate recipients. In the last two decades, many different recommendation algorithms have been developed [39]. Basically, the initial step of our profiling approach imitates recommendation based on item similarity [40]. In our case, the item is a musician, and the system determines similar musicians as recommendations. The further profiling steps of our systems are rather comparable to content-based recommendation systems [42]. An item is recommended based upon description and the "profile of the user's interests", which are user-defined similarity weights and mandatory musician attributes in our case. The benefit of using weights to reflect the importance of certain attributes for users has already been shown for discovering related movies [18]. Other application areas include the development of recommender systems for YouTube videos [17], music files [11], or related people in social networks [12]. The recommendation of social media data proposed by Guy et al. [28] is based on relationship information among people. Our system uses this idea and determines a relationship similarity between musicians.

2.2 Related Visual Analytics & InfoVis Techniques

The visualization of relational, geospatial, temporal and nominal textual data is a common task for researchers in visual analytics and information visualization. In accordance to the features of our profiling system, we outline the most related works under various aspects.

Visualizing recommendations Many works provide visualizations for recommendations of different type. Choo et al. present a recommendation system for a vast collection of academic papers [13]. Recommended documents can be explored in a scatterplot. Overview is a system that also supports the systematic analysis of large document collections to be used by investigative journalists [8]. Documents are hierarchically clustered based on content similarity, thus, can be recommended during investigation. A recommendation of notes from past analysis tasks when operating a visual analytics system is explained by Shrinivasan et al. [47]. For the visualization of recommended movies, Vlachos uses a radial layout [50]. Gansner offers geographic maps that use the metaphor of neighborhoods and clusters to group related recommendations [23]. The latter approach communicates relations among recommendations. That a recommendation system is not seen as a black box from the user's point of view [49], we provide visual and textual cues for the collaborating musicologists to explain the existence of a recommendation.

Social network visualizations In our profiling system, relationship graphs visualize the social networks of musicians under inspection. Especially interesting are unknown musicians that connect unacquainted musicians. Applied to the relationships of characters in literary works, Euler diagrams can be used to visualize groups of social networks in clusters while also showing relations among clusters [44]. Weaver embeds an attribute relationship graph in his visual analytics environment to visualize the relationships between movie actors [54]. Quite often, visualizations attend to the matter of visualizing large social networks and their navigation by various interaction means [43]. The design of such visualizations is especially important for the exploration of online social networks [27, 31].

Multiple views Our system consists of multiple views that visualize the musician's metadata information, e.g., we provide visual representations of temporal, geospatial, and nominal textual data. The multiple views concept is often used to communicate various data aspects. An overview of various visual analytics approaches to dynamically explore spatio-temporal data with the help of maps and timelines is given by Andrienko et al. [4]. The purpose of such visual analytics environments ranges from the analysis of crime incidents [33] to the extraction and characterization of significant places from mobility data [3], the visualization of semantic web data [10], and the comparative analysis of geospatial-temporal data [36]. Also popular is the additional visualization of contextual keywords – next to map and timeline – in the form of tag clouds, for instance, to visualize topical metadata [20] or to support the discovery of meaningful events in news and social media data [22]. Many approaches are based on textual data sources. The cross-filtered views for multidimensional data sets as proposed by Weaver include various interfaces to be used as filters to determine potentially interesting events in newspaper article collections [53]. Heimerl et al.'s visual analytics system is also composed of many views with the purpose of interactively training classifiers to be used for document retrieval on large text collections [32].

Layered Textual Metadata One of the major components of our system is the Column Explorer (see Section 5.2) with a column per data facet showing textual attributes of musicians. Jigsaw's list view [48] provides the basic idea for such a visualization. The user can select an arbitrary entry and correlations to other attributes are shown. Related data entries in adjacent columns are linked. Similarly, this is done in Parallel Tag Clouds [14]. Each column lists tags of a certain time slice, and equal tags are linked on selection. PivotPaths is a yet similar approach to ours [21]. After selecting a research paper, links are drawn to related authors and paper keywords. The related attributes in our Column Explorer are connected with colored streams. Often, streams express a temporal evolution of events [9, 16]. Tags at certain positions in streams can be used to illustrate contextual information [46].

2.3 Visualizations for the Digital Humanities

The database of our visual profiling approach is based upon the textual contents of numerous digitized documents about music history. Many visualization techniques with a motivation from the humanities also provide abstract views on digital text editions. Most often, the focus of interest is a specific literary text.

Keim et al. present fingerprint matrices that visualize extracted features, which characterize a given text [37]. The purpose is to support the analysis of the behavior of feature values across the text. Another approach concerns the visualization of poem features. Proposed by Abdul-Rahman et al. [2], the Poem Viewer uses visual attributes to encode phonetic units as well as phonetic and semantic relationships. VarifocalReader facilitates the work with individual, potentially large digital texts by providing visualizations on various text hierarchy levels [38]: an overview of the text structure, tag cloud summaries, and a close reading view. Applied machine learning techniques and search mechanisms support the user to extract entities, concepts and other artifacts from the examined text. Correll et al. provide a text analysis environment for a whole corpus [15]. It aims to allow the detection of corpus-wide statistical patterns, and texts can be displayed the way that text passages reflect the relevance according to the user's preferences. Vuillemot et al. provide a flexible system for the dynamic exploration of a single literary text [51]: *The Making of Americans* by Gertrude Stein. Various interfaces like tag clouds and self-organizing graphs support to review vocabulary, to filter by part of speech, and to explore character networks. In GeneaQuilts [7], the genealogy of extracted characters from a literary text such as the Bible are visualized.

The purpose of our work is to visualize the similarity between musicians. Some related works on issues in literary criticism visualize the similarity between various text passages. The Word Tree shows all sentences of a given text that share the same beginning in the form of a tree [52]. Other approaches highlight differences and similarities among various editions of a text, e.g., various German translations of Shakespeare's Othello [24] or various English translations of the Bible [35].

3 Digital Humanities Background

This research bears on musicology, a field of the humanities that observes musicians and their achievements. This includes not only composers, although a *composition* is seen as the fruit of a musical process. Moreover, many other musical professions are in the focus of interest of musicological research, e.g., instrument makers, conductors, singers, instrumentalists, music publishers, etc.

Motivation The Bavarian Musicians Encyclopedia Online (BMLO) project was initiated in 2004 with the goal to create a database that combines a multitude of biographical information about musicians of various professions. In cooperation with the Bavarian State Library and the Society for Bavarian Music History, musicologists of the Ludwig Maximilian University of Munich searched, collected and digitized related documents on music history. They combined biographical information about musicians extracted from various sources such as encyclopedias, periodicals, and series concerning musicology as well as research papers from musicology, history and science of art. A web-based platform [1] provides access to the database, which contains musicians who are part of the Bavarian music history; musicians with an active lifetime period living in Bavaria as well as musicians with a considerable influence on the Bavarian music history are included. Despite the prior focus on Bavaria, the BMLO is a valuable tool for many musicologists as it provides information about 28,137 musicians from all musical eras, spanning a time range from 4AD to the present. Working with the BMLO, the main interests are not examinations of the musicians' achievements – musicologists rather explore the features of musical professions or analyze the biographies of musicians. This includes generic research questions concerning the geographical or temporal evolution of musical professions as well as precise research questions that focus on an individual musician. One such research question – the profiling of musicians with similar careers to a musician of interest – is interesting for musicologists for a long time. Traditionally approaching an answer to this type of question, musicologists solely refer to musicological editions and monographs. But musicology primarily focuses on fifty musicians – mostly composers – and their main works. Due to this inhomogeneous state of research (what we call the popularity of musicians), a traditional similarity analysis usually starts and ends within this limited set of musicians. Although the BMLO provides an immense diversity of information about a large number of musicians, the profiling of musicians is not supported. Maybe the database could be used to address some research questions, but for musicologists complex database queries are hard to formulate [25] and the musicians' attributes in the query result are hard to analyze and to compare. Therefore, musicologists desired a system that allows to approach a profiling task interactively with the aid of visual interfaces that pictorially illustrate the provided information.

Digital Humanities Project In close collaboration with musicologists using the BMLO, we developed a visual analytics system that supports the profiling of similar musicians based on a selected musician of interest. For the implementation of this project, we adopted several suggestions made by Munzner [41] to ensure designing a beneficial, powerful tool that supports answering the posed research questions. We furthermore took collaboration experiences [34] from other visualization researchers who worked together with humanities scholars into account to avoid typical pitfalls of such interdisciplinary projects. Additionally, we worked through related works in the digital humanities, which provide valuable suggestions and guidelines for designing interfaces for humanities scholars, e.g., outlined in [25]. To avoid making assumptions for the design of a profiling system that is hard to comprehend and does not solve the concerned musicological research questions, we initially discussed the needs of the musicologists, their workflows and challenges in the targeted domain in several meetings. Furthermore, we presented and discussed related visualization techniques to convey an impression of the capabilities and challenges within our research field. The musicologists explained how they use the BMLO for their workflows and communicated their fascination about this unique type of encyclopedia invaluable for their daily work. This get together turned out to be important to understand each others mindsets. A major outcome was a set of research questions on the profiling of musicians and the analysis of musician profiles.

Project Data The provided database and aspects of data transformation were also discussed with the musicologists. This included occurring data anomalies, the conversion of the temporal metadata to a uniform scheme while considering occurring uncertainties as well as defining popularity values by counting a musician's references. In discussions about the provided musician attributes we could separate attributes worth to integrate into the profiling process – a musician's sex, lifetime data, places of activity, musical and further professions, relationships, divisions and denomination – from irrelevant ones. For instance, the potentially interesting attribute *nationality* is only provided for 416 musicians (1.5%) as most musicians lived in a time when the assignment of a nationality to a person did not exist. Therefore, we decided to exclude nationalities from the profiling process. The musicologists argued that most research interests concern musicians without nationality attributes. We also asked for the relevance of each attribute dimension for a profiling task and the comparative analysis of musicians in order to push the development of the profiling system the way that predominant attributes receive more attention. Additionally, we gained information how musicologists imagined to operate with the musicians' attributes. For example, they wanted to see how attributes of different facets correlate, and they wanted to detect the links between unrelated musicians.

Project Challenges To solve the profiling task, we faced two main challenges. On the basis of relevant musician attributes, we first needed to define various similarity measures that determine the similarity of musicians (see Section 4). Second, we needed to design visual interfaces that communicate these similarities intuitively. In preparation, we looked at related visualizations and collected possible representations to map relevant attributes of musicians to visual attributes. In meetings with the musicologists, we argued on opportunities and drawbacks when applying various visualization techniques. The resultant visualization design is explained in Section 5. Finally, various usage scenarios illustrate the utility of the profiling system for the collaborating musicologists (see Section 6), now capable of detecting similar musicians without the bias of popularity. As further demands included the visual exploration of individual musician profiles and the comparative analysis of multiple profiles, we also provide an example besides profiling.

uncertainty	dating year	difference
before/after	≤ 1700	$-/+30$ years
	$1701-1800$	$-/+25$ years
	$1801-1900$	$-/+10$ years
	>1900	$-/+5$ years
around	≤ 1500	± 20 years
	$1501-1600$	± 15 years
	$1601-1700$	± 8 years
	$1701-1800$	± 5 years
	$1801-1900$	± 3 years
	>1900	± 2 years

Table 1. Mapping of uncertain datings.

4 THE SIMILARITY OF MUSICIANS

Based on various biographical information, the similarity $S(m_i, m_j)$ between the musician of interest m_i and a similarity candidate m_j is determined as

$$S(m_i, m_j) = w_p \cdot P(m_j) + \sum_{k=1}^{8} w_k \cdot S_k.$$

w_k is a weight for the relevance of the corresponding similarity S_k. To mimic the traditional profiling approach by referring to musicological editions and monographs, we insert the popularity $P(m_j)$ of m_j as a further component into the similarity equation. The collaborating musicologists define $P(m_j)$ in dependency on the number of publications (articles, editions and media) from and about m_j. Thus, the popularity reflects the current state of research on a musician. According to this heuristic, Wolfgang Amadeus Mozart is the most popular musician with around 150,000 publications. Taking the musicologists' suggestions into account, we group all musicians with the same number of publications into groups $g_1, \ldots, g_n$ sorted by ascending publication count. The popularity $P(m_j)$ is then defined in dependency on m_j's popularity group g_k as

$$P(m_j) = \frac{k}{n}.$$

w_p can be used to adjust the influence of popularity during the profiling process. Using $w_p = 0$ disregards popularities and $w_p = 1$ mimics the traditional profiling approach. All weight values are defined interactively during the profiling process. In the following, we outline the calculation for each of the eight contributing similarities. Some of the similarity measures are defined by the Jaccard index like

$$S_i(m_i, m_j) = J\Big(f(m_i), f(m_j)\Big) = \frac{|f(m_i) \cap f(m_j)|}{|f(m_i) \cup f(m_j)|}.$$

4.1 Sex Similarity S_1^{sex}

For some research questions of the collaborating musicologists, the sex of a musician plays an important role when determining similar musicians. Such an information in the form of *male* or *female* is provided for nearly all musicians (27,403 $\hat{=}$ 97.4%). If the sexes of m_i and m_j are equal, we define $S_1^{sex}(m_i, m_j) = 1$. For unequal sexes or if the sex of one musician is unknown, we use $S_1^{sex}(m_i, m_j) = 0$.

4.2 Activity Time Similarity S_2^{tem}

The activity time of a musician is defined in dependency on the temporal metadata provided for nearly all musicians of the database (27,681 $\hat{=}$ 98.4%). Three various datings may be given for a musician: a dating of birth B (provided for 27,357 musicians $\hat{=}$ 97.2%), a first mentioned dating F (25,592 $\hat{=}$ 91%), and/or a dating of death D (18,610 $\hat{=}$ 66.1%)

The musicologists exploited the underlying textual sources of the database the way that the first mentioned dating is always an evidence for an active phase of a musician, thus, always ranges between birth and death. The granularity of the given datings ranges from date to year. Due to uncertain information in the textual sources, the given datings are often imprecise. Three types of uncertainty occur: *before* datings (e.g., before 1745), *around* datings (e.g., around March, 1745), and *after* datings (e.g., after September 22, 1745).

In order to process uncertain datings for the purpose of defining and visualizing the activity time for each musician, the collaborating musicologists provided a taxonomy – based on state-of-the-art knowledge in musicology – to map uncertainties to years as approximate datings. Table 1 lists how various uncertain datings are resolved in dependency on centuries. For all *before* and *after* datings we add or subtract the given difference value. For datings with an *around* uncertainty, we subtract the difference value for births, add the difference for deaths, and for first mentioned datings we define F_{min} by subtracting and F_{max} by adding the difference value. In few cases, irregularities occur after resolving uncertain datings. In case of $F_{min} < B$ (or $F < B$) we set $F_{min} = B$ (or $F = B$), and if $F_{max} > D$ (or $F > D$) we set $F_{max} = D$ (or $F = D$).

The activity time $t(m) = \{t_{min}(m), t_{max}(m)\}$ of a musician m is determined based upon the given dates as follows:

- **if** F or F_{min} and D are defined and unequal, we set $t_{min}(m) = F$ or $t_{min}(m) = F_{min}$ and $t_{max}(m) = D$
- **else if** F_{min} and F_{max} are defined, we set $t_{min}(m) = F_{min}$ and $t_{max}(m) = F_{max}$
- **else if** F is provided, we define F_{max} by applying the *after* uncertainty to F and use $t_{min}(m) = F$ and $t_{max}(m) = F_{max}$
- **else if** B and D are provided, we use $t_{min}(m) = B + 20$ *years* and $t_{max}(m) = D$

In the rare cases if only B or only D are provided, the definition of an activity time range is too hypothetical according to the musicologists. In such cases, the corresponding similarity is always $S_2^{tem}(m_i, m_j) = 0$. In case of two valid activity time ranges, we define $S_2^{tem}(m_i, m_j)$ using the Jaccard index as $S_2^{tem}(m_i, m_j) = J\big(t(m_i), t(m_j)\big)$.

4.3 Activity Region Similarity S_3^{reg}

The database contains places of activity, where musicians lived or worked for a certain period of time. At least one such place is provided for 26,101 musicians (92.8%) in the database. For the most often occurring 1,661 places of activity, geographical coordinates as longitude/latitude pairs and hierarchical place IDs for the contemporary political belonging of a place are given.

The activity region of a musician consists of all places of activity. The similarity $S_3^{reg}(m_i, m_j)$ between the activity regions of m_i and m_j is determined taking the political belongings as well as the geographical positions of the musicians associated places into account. For this

place	id	hierarchy levels	1.	2.	3.	4.
1. Bonn	XA–DE–05–3–14	Europe–Germany–North Rhine-Westphalia–Cologne (county)–Bonn	1.0	0.4	0.4	0.4
2. Munich	XA–DE–09–1–62	Europe–Germany–Bavaria–Upper Bavaria–Munich	0.4	1.0	0.6	0.8
3. Nuremberg	XA–DE–09–5–64	Europe–Germany–Bavaria–Middle Franconia–Nuremberg	0.4	0.6	1.0	0.6
4. Erding	XA–DE–09–1–77–117	Europe–Germany–Bavaria–Upper Bavaria–Erding (county)–Erding	0.4	0.8	0.6	1.0

Table 2. Political identifiers of four German cities and their political distances.

purpose, we define the two measures Political Distance D_{pol} and Geographical Distance D_{geo}.

Political Distance D_{pol} The (contemporary) political distance $D_{pol}(p_1, p_2)$ between two places p_1 and p_2 is defined in dependency on hierarchical place identifiers provided for each place. The level of detail of such an identifier varies from one (only continent) to seven. Examples are listed in Table 2. For most places, at least five hierarchy levels are given. Therefore, we define $D_{pol}(p_1, p_2)$ dependent on k first equal hierarchy levels as

$$D_{pol}(p_1, p_2) = \frac{k}{5}.$$

Geographical Distance D_{geo} To determine the geographical distance $D_{geo}(p_1, p_2)$ between two places $p_1 = \{x_1, y_1\}$ and $p_2 = \{x_2, y_2\}$ in kilometers, we use the great circle distance G, taken from [30]:

$$G = 6378 \cdot \arccos\Big(\sin(y_1)\cdot\sin(y_2) + \cos(y_1)\cdot\cos(y_2)\cdot\cos(x_1 - x_2)\Big).$$

$D_{geo}(p_1, p_2)$ is then defined as

$$D_{geo}(p_1, p_2) = \frac{d_{max} - G}{d_{max}}.$$

Specified by the musicologists, d_{max} is the maximum distance allowed for two places to be geographically related in former times. For the examples in this paper, we used $d_{max} = 500$km – empirically determined by the musicologists. In case of $G > d_{max}$, we define $D_{geo} = 0$.

Given two sets P_i and P_j of places of activity for m_i and m_j, we use the iterative closest point algorithm [6] to calculate the activity region similarity $S_3^{reg}(m_i, m_j)$. For each place p_i^k in P_i, we determine the distance $d(p_i^k)$ to the "closest place" in P_j, which we define as

$$d(p_i^k) = \max_{p_j^l \in P_j} \Big(D_{pol}(p_i^k, p_j^l)\cdot D_{geo}(p_i^k, p_j^l)\Big).$$

Likewise, we determine the distance $d(p_j^l)$ to the "closest place" in P_i for each place $p_j \in P_j$. Finally, $S_3^{reg}(m_i, m_j)$ is defined as

$$S_3^{reg}(m_i, m_j) = \frac{\sum_{k=0}^{|P_i|} d(p_i^k) + \sum_{l=0}^{|P_j|} d(p_j^l)}{|P_i| + |P_j|}.$$

4.4 Musical Profession Similarity S_4^{mus}

For 26,695 musicians (94.9%), the database contains information about their musical professions such as composer, conductor or pianist. Musical professions are of special importance for the musicologists as they substantially define the emphasis of a musician's activity. They are given as lists $mus(m)$ for each musician m, and the similarity $S_4^{mus}(m_i, m_j)$ between the musical professions of m_i and m_j is defined by the Jaccard index as $S_4^{mus}(m_i, m_j) = J\big(mus(m_i), mus(m_j)\big)$. In case of $|mus(m_i)| = |mus(m_j)| = 0$, we define $S_4^{mus}(m_i, m_j) = 0$.

4.5 Further Profession Similarity S_5^{pro}

The database also provides information about professions unrelated to music (e.g., philosopher, teacher, soldier) for 7,920 musicians (28.1%). As above, the Jaccard index is used to determine the similarity $S_5^{pro}(m_i, m_j)$ for the further professions $pro(m_i)$ and $pro(m_j)$ of m_i and m_j as $S_5^{pro}(m_i, m_j) = J\big(pro(m_i), pro(m_j)\big)$. In case of $|pro(m_i)| = |pro(m_j)| = 0$, we define $S_5^{pro}(m_i, m_j) = 0$.

4.6 Relationship Similarity S_6^{rel}

One of the key features of the database are the inherent social relationships. For the collaborating musicologists, these information generate an invaluable social network that reflects interpersonal relationships of the most important musicians in the musical landscape, although relationships are only provided for 9,739 musicians at the moment (34.6%). Nevertheless, the resultant social network contains large communities as connected components. The largest community is composed of 5,065 musicians. Each relationship has a specific type and a role is assigned to both connected musicians. The musicologists also defined the strength s_{rel} for each relationship type (see Table 3). The distance between two related musicians m_k and m_l is defined in dependency on the relationship strength $s_{rel}(m_k, m_l)$ as:

category	relationship	s_{rel}
family of origin	parents, children, siblings grandparents, grandchildren	1
partnership	partners	1
education	fellow students, teachers, students	0.8
relatives	cousins, nephews, nieces, uncles, aunts, great uncles, great aunts, grandnephews, grandnieces	0.6
godparenthood	godparents, godchildren	0.6
affinity	parents in law, children in law brothers/sisters in law	0.4
personal relationships	network, patrons, protégés	0.4
working environment	colleagues, predecessors, successors	0.2
dedication	dedication donors & recipients	0.2

Table 3. Relationships and their strength s_{rel}.

$$d(m_k, m_l) = \frac{1}{s_{rel}(m_k, m_l)}.$$

Taking all relationships of the database into account, the relational similarity $S_6^{rel}(m_i, m_j)$ of m_i and m_j is derived from the shortest path $p(m_i, m_j) = \{m_i, \ldots, m_j\}$ connecting both musicians in the social network graph and its length $|p(m_i, m_j)|$, determined using Dijkstra's algorithm [19]:

$$S_6^{rel}(m_i, m_j) = \sum_{k=0}^{|p(m_i, m_j)|-1} \frac{1}{k+1}\cdot d(p[k], p[k+1]).$$

Thus, the similarity between acquainted musicians is $S_6^{rel}(m_i, m_j) = s_{rel}(m_i, m_j)$ and the similarity for musicians unconnected in the graph is $S_6^{rel}(m_i, m_j) = 0$.

4.7 Division Similarity S_7^{div}

Further important characteristics are the divisions where musicians worked (e.g., court, theater). These information are given for 17,062 musicians (60.6%). We determine the similarity $S_7^{div}(m_i, m_j)$ between the known divisions $div(m_i)$ and $div(m_j)$ of the musicians m_i and m_j using the Jaccard index: $S_7^{div}(m_i, m_j) = J\big(div(m_i), div(m_j)\big)$. In case of unknown divisions $|div(m_i)| = |div(m_j)| = 0$, we define $S_7^{div}(m_i, m_j) = 0$.

4.8 Denomination Similarity S_8^{den}

Especially in former times, the denomination(s) of a musician influenced her activity in a particular manner. Although this information is not provided for 21,302 musicians of the database (75.7%), research questions may include references to a musician's denomination(s). One denomination is given for 6,733 musicians, two denominations for 117 musicians, and for two musicians even three denominations are provided. Therefore, we use the Jaccard index also to determine the denomination similarity $S_8^{den}(m_i, m_j)$ between m_i and m_j in dependency on the musicians' denominations $den(m_i)$ and $den(m_j)$ as $S_8^{den}(m_i, m_j) = J\big(den(m_i), den(m_j)\big)$. In case of unknown denominations $|den(m_i)| = |den(m_j)| = 0$, we define $S_8^{den}(m_i, m_j) = 0$.

5 The Profiling of Musicians

The idea of musician profiling is to detect a user-defined number N of similar musicians $s_1,\ldots,s_N$, who shared similar attributes with a given musician m of interest. The profiles of all observed musicians are visualized in three different views: Column Explorer, Relationship Graph and Map.

5.1 Profiling Workflow

Initially, the musicologist enters the musician m of interest for whom the profile is visualized in the previously mentioned visual interfaces. Observing the various attributes of m, the scholar is able to define mandatory profiling attributes. A similar musician s then requires to share this attribute. Possible mandatory attributes are:

- **Musical & further professions, divisions, denomination(s)**: s shares all mandatory attributes of m in these categories.
- **Activity time**: The intersection of the activity time ranges of m and s is not empty.
- **Place(s) of Activity**: All mandatory places of activity of m were also places of activity of s.

The selection of mandatory attributes supports specific research questions like "Find the most similar musicians to Wolfgang Amadeus Mozart with the musical profession *concertmaster* who worked at a *court* and who had *Salzburg* as place of activity!" Mozart worked as a concertmaster at the court of Salzburg between 1772 and 1777. Although the database does not contain information if a musician worked at a specific place in a certain profession, the system is capable of providing hints to investigate such questions.

After selecting mandatory attributes, the musicologist performs the first profiling iteration based on all similarity measures defined in the previous section. The weight w_i of a similarity measure S_i is automatically determined in dependency on the diversity of available attributes in relation to the attributes of m in the corresponding dimension. With the number n of musicians $m_1,\ldots,m_n$ with the given attribute ($m \notin m_1,\ldots,m_n$), we define w_i as

$$w_i = 1 - \frac{\sum_{k=1}^{n} S_i(m,m_k)}{n}.$$

An example is given by the weight w_1 for sex similarity S_1^{sex}. The database contains 23,865 male musicians (84.8%), 3,538 female musicians (12.6%) and 751 musicians (2.7%) without a sex information. When profiling similar musicians for Wolfgang Amadeus Mozart, the initial weight for sex similarity is $w_1 = 0.13$ as the database contains mostly male musicians. An initial profiling on Wolfgang Amadeus Mozart's wife Constanze Mozart would use $w_1 = 0.87$ due to the comparatively small number of female musicians.

The result of the first profiling iteration are N similar musicians $s_1,\ldots,s_N$. As outlined above, the profiles of $s_1,\ldots,s_N$ are visualized alongside the profile of m. That individual attributes are easy to track, a certain color is assigned to each musician. As N is usually small – less than ten similar musicians –, we use the ColorBrewer [29] to generate a qualitative color map that provides solely saturated colors to be used on the bright website background. In further iterations of the profiling process, the musicologist can gradually modify mandatory attributes and similarity weights as desired in order to receive similar musicians with certain attributes relevant to the posed research question. In Section 6, we illustrate several usage scenarios with interesting findings to emphasize the benefit of this interactive visual analytics approach for the collaborating musicologists.

5.2 Column Explorer

Inspired by Jigsaw's list view [48] and Parallel Tag Clouds [14], we designed an interface that allows for the exploration of various metadata information provided. The Column Explorer consists of various columns that serve various purposes.

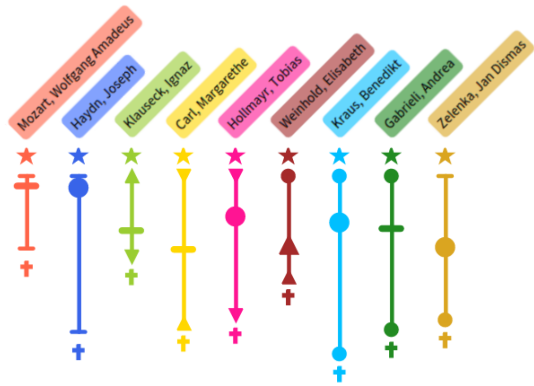

Fig. 2. Lifetime data examples: various shapes encode uncertain birth, death and first mentioned datings.

Legend All observed musicians are shown in the form of a legend in the leftmost column. m is positioned at the top, and $s_1,\ldots,s_N$ are listed below, ordered by descending similarity to m. The background of a musician's name is drawn in the musician's assigned color. Hovering a musician lists the following attributes in a popup: sex, popularity rank, nationalities and BMLO identifier.

Lifetime Data In a vertical timeline, the temporal metadata of all observed musicians is visualized in vertical slots. If provided, we put marks for the date of birth (additionally highlighted with a star symbol ★), the first mentioned date (slightly larger mark), and the date of death (additionally highlighted with a cross symbol †). The shapes used as marks reflect the precision of the provided dating. A small horizontal line ▬ is used for precise datings, and circles ● highlight *around* datings. Triangles ▲ mark *before* datings as they point to the start of the vertical timeline, thus, upside down triangles ▼ illustrate *after* datings. The lifetime of a musician is shown with a vertical line in the corresponding color that connects dates of birth and death. Various examples are shown in Figure 2.

Nominal Textual Metadata Four columns list the occurring musical and further professions, divisions and denomination(s) of m and $s_1,\ldots,s_N$. In each column, the attributes of m are listed first in alphabetical order. By descending similarity, further attributes of the determined similar musicians are listed. Being the most powerful metaphor of tag clouds [5], we use variable font size of labels to encode the number of attribute occurrences. If a musician does not have an attribute in a certain column, we put a "no information" label to communicate this information – an often mentioned demand of the musicologists of our project. Clicking an attribute label toggles its mandatory selection for the profiling process. Only attribute labels (except "no information" labels) belonging to m can be selected.

In Jigsaw's list view, attributes from different columns are connected if they belong to the same data entity. Also, in Parallel Tag Clouds various tags are connected after selection. We use these ideas and display the coherence of the attributes of each musician as colored streams passing all columns of the Column Explorer. Starting from the legend, a stream marks the activity time of the corresponding musician in the timeline. Thereby, the occlusion of streams illustrates similar activity times. After passing the timeline, a stream runs through all related attributes of a musician. In case of multiple attributes, a stream splits and passes the corresponding attributes. This metaphor aims to visualize occurring correlations among various attributes of musicians and to further facilitate the visual comparison of different profiles.

5.3 Relationship Graph

The relationship graph of our profiling system is invaluable for the collaborating musicologists as it provides the view on a musician's social network for the first time visually. Furthermore, musicians that connect two observed musicians of interest become visible. For many musicians, a list of relationships to other musicians in the database

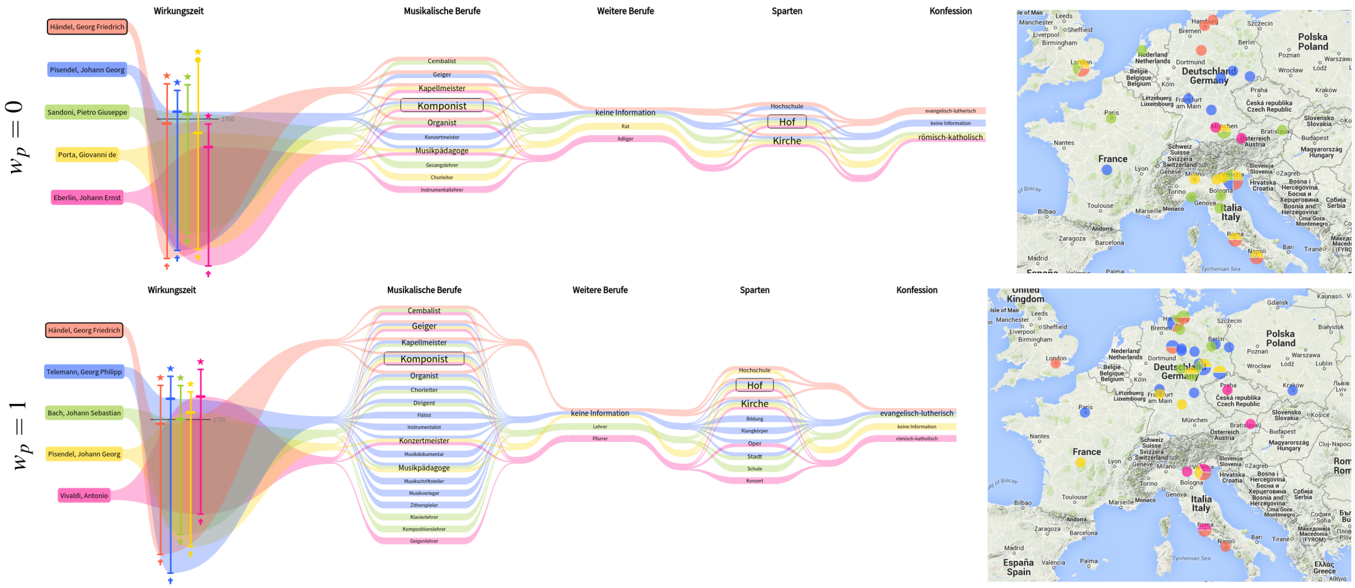

Fig. 3. A profiling for musicians similar to court composer Georg Friedrich Handel. The musical profession *composer* (Komponist) and the division *court* (Hof) are marked as mandatory. The results are shown disregarding popularity ($w_p = 0$) and taking popularity into account ($w_p = 1$).

is provided. Possible relation types between two musicians and the strength of each relationship are shown in Table 3. Taking all relationships of m and $s_1, \ldots, s_N$ forms a social network graph with vertices representing musicians and edges connecting related musicians. In order to facilitate an easy exploration of the graph, we only take the direct relationships of each musician into account. Furthermore, we add the relationships between related musicians $\notin \{m, s_1, \ldots, s_N\}$ to receive a closed social network. We use a force-directed algorithm to generate the network graph. We thereby map the strengths of relationships to intended ideal edge lengths when computing the layout. An example social network is shown in Figure 4. The observed musician vertices for m and $s_1, \ldots, s_N$ are drawn in the corresponding color, and their full name is shown next to it. To keep the social network explorable, all additional musicians are drawn as gray vertices, and only the first four letters of their names are shown. The latter design decision reduces the occurrences of occluding labels to a minimum while alongside providing an "adequate information" for the musicologist, who is usually aware of the social relations of the observed musician(s). More detailed information can be shown using mouseover interaction. Hovering a gray vertex pops up the full name of the corresponding musician, whereas hovering an edge provides the roles of the two connected musicians in their relationship. Unobserved, but potentially interesting musicians shown in the social network graph can be added to the profile visualization via mouse click. A mouse click onto the vertices representing m and $s_1, \ldots, s_N$ visualizes the shortest paths to all other musicians under investigation. This feature supports the musicologists in examining the channels through which information was most probably transferred in former times.

5.4 Map

The map of the profiling system visualizes all places of activity provided for m and $s_1, \ldots, s_N$. The focus of interest is to facilitate the visual interpretation of a certain activity region and to support the comparison of different activity regions. A location that was only the activity place of one of the musicians under inspection is displayed as a single circle with a radius r_c drawn in the corresponding color. Quite often, musicians shared the same places of activity. For example, Munich was a place of activity for 10,558 musicians of the database (37.5%). To forestall the misinterpretation of activity regions through occluding individual circles, we draw a pie chart for each shared place. We scale the radius r_p of a pie chart dependent on the number of associated musicians to avoid visual distortion. To receive pie slices with the same area as an individual circle, we define r_p as

$$r_p = \sqrt{N} \cdot r_c.$$

All shapes are drawn slightly transparent to avoid losing the geographical context in dense regions. An example is given in Figure 5. Hovering a shape displays a popup that shows the place name and a list of related musicians.

6 Usage Scenarios

The traditional approach of searching for similarities between musicians is biased due to the inhomogeneous state of research (popularity). Rather observing the similarities among the other musician's attributes, the usage of the profiling system revealed substantial anomalies in contrast to this traditional approach. The first out of four scenarios provided by musicologists using our system exemplifies this issue.

Profiling Georg Friedrich Handel Handel is one of the most popular court composers. A musicologist used the profiling system to iteratively discover court composers of the same era (mandatory activity time, division *court* and musical profession *composer*) with similar careers. The initial profiling result shows similar musicians from different generations with first mentioned datings ranging from 1691 to 1731. Now increasing the weight for activity time ($w_2^{tem} = 1$) and ignoring activity regions ($w_3^{geo} = 0$) better models the musicologists imagination of a "same era" by narrowing this time range (1691-1708). Then, the musicologist tests various combinations of weights for popularity and denomination disregarding relations ($w_4^{rel} = 0$) with interesting insights (in all combinations the era range remains small). The set of similar musicians for various popularity settings and $w_8^{den} = 1$ changes only slightly and always contains popular Evangelical-Lutheran musicians like Johann Sebastian Bach and Georg Philipp Telemann. By further applying varying denominational significance and using $w_p = 0$, the musicologist discovers unexpected, very similar profiles to Handel in terms of musical professions and divisions for rather unknown musicians with activity places in southern European regions; especially, two Italian musicians are in the result set. With Venice, Rome and Naples, Handel had an active period in three Italian cities. As $w_3^{geo} = 0$ was used, this correlation hypothesizes mutual influences between Handel and the musicians found as well as an Italian influence on Handel's work. Now mimicking the traditional profiling approach based upon print media by applying

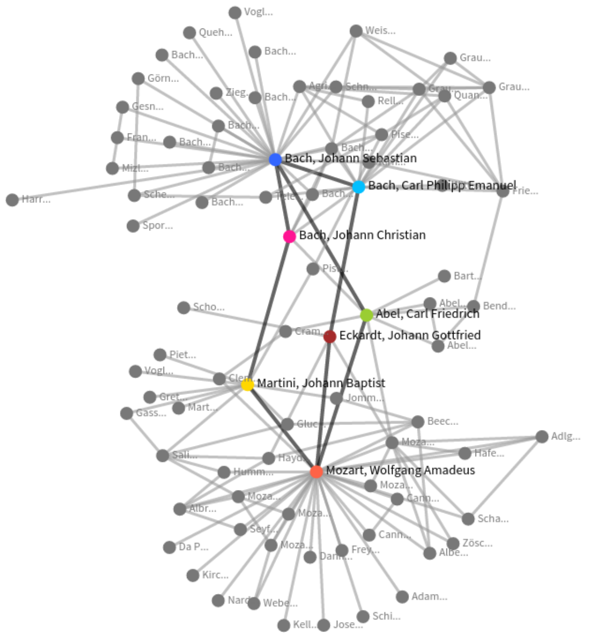

Fig. 4. The social network shows potential pathways how Bach's score was transferred to Mozart (related links are highlighted for illustration).

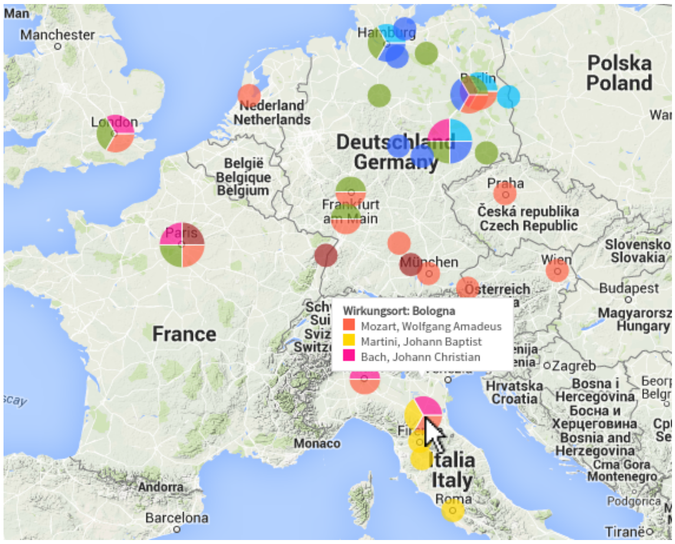

Fig. 5. The map shows places of activity from Mozart, Bach, and relationships who connect both musicians in the social network to analyze potential places where related musicians met each other or worked together. The city Bologna marks a shared place of activity on one possible transfer path of Bach's score to Mozart.

$w_p = 1$, most of the rather unknown musicians are replaced by popular ones with a lesser similarity regarding musical professions and divisions (see Figure 3). Especially the similarity between Handel and Antonio Vivaldi – both sharing only few characteristics – seemed accidental to the musicologist. Thus, this use case exemplifies the biased influence of popularity. But the tool opens new research perspectives by focusing rather unknown but more similar musicians as opposed to focusing popular musicians.

Profiling Meinrad Spieß Starting a profiling for musicians similar to monastery composer Spieß, a musicologist would predominantly refer to musicological editions and monographs. According to them, the results of this traditional approach would be again biased due to the inhomogeneous state of research (popularity). Our profiling system was used to search for musicians similar to Spieß disregarding popularity ($w_p = 0$). The initial profiling step shows a list of other southern German Catholic church musicians of the early modern era. To further specify the profile scheme, the musicologist increases the weight for activity region, activity time, division and denomination similarity, whereas the weight for relationship similarity is lowered. As the result differs only slightly, the weights for denomination and activity region are set to 0. Then, middle German Protestant contemporaries occur – a comprehensible fact as Spieß (1) was an active member of Lorenz Mizler's musical circle in Leipzig, and (2) corresponded frequently with academy colleagues outside his denominational bounds and activity region. A particular observation is an obvious similarity to musicians belonging to a generation that essentially characterized the *bandmaster* profession. Similarities are discovered to the known musicians Johann Sebastian Bach and Georg Friedrich Handel, and other representatives of this first bandmaster generation – a fact hardly recognized in previous music research. Approaching insights this way is a novel technique in musicology. An also known, but never visualized phenomenon in musicology are homogeneous subcultures of the early modern era. By only regarding relational similarities for Spieß, the musicologist detects a closed Benedictine network composed of Spieß' students and the relatives of his own teacher Bernabei.

Profiling Maier Kohn The cantor is one of the most multifarious musical professions in cultural history. The characteristics of this profession equally depend on chronology, cultural area and denomination. Consequently, the responsibilities of a cantor are widespread and the term *cantor* is impractical to be used as musical vocabulary. Subsequently, cantors cannot be easily retrieved using the Bavarian Musicians Encyclopedia Online. The musicologist requires an analysis in dependency on the various duties of the cantor profession. Using the profiling system, this multifacetedness can be visualized and analyzed in individual cases. The Jewish cantor Maier Kohn had many musical responsibilities in clerical music and in school, as a singer, organist, composer and choirmaster. Similar profiles regarding musical professions that compose the cantor profession can be found in various contexts for musicians of different generations, with different activity regions and denominations. Limiting the profiling to Jewish musicians with similar profiles provides a list of Jewish cantors (e.g., Salomon Sulzer and Joseph Freudenthal) with many varying musical professions (see Figure 1). This brief meta-analytic test reveals a contradiction to the anti-Semitic influenced state of research of the early 20th century. At that time, musicologists claimed a monotonous interpretation of the cantor profession for Jewish musicians. But the multifaceted musical professions of all cantors in the result suggest a diversity similar to known Christian cantors. According to the anti-Semitic research, Maier Kohn was therefore not a "typical Jew". His strong similarity to Christian cantors when disregarding Jewish denomination in the profiling underpins this fact. This example outlines the utility of the profiling system to transform a musicological issue – the multifacetedness of the musical profession cantor –, which is not existent in the database, into a representation visualized as the Column Explorer.

The missing link between Johann Sebastian Bach & Wolfgang Amadeus Mozart This example illustrates the usage of the system without its profiling capacity. Mozart was born few years after Bach's death, but Mozart played Bach's music. Mozart primarily worked at southern German Catholic residences where the trade with music supplies was unincisive. An interesting research question for the musicologist arose: "What was the connection through which Bach's score was transferred to Mozart?" First, the musicologist visualizes the profile of both musicians. Second, the relationship graph is explored and candidates on probable pathways between Bach and Mozart are added to the profile visualization (see Figure 4). Taking all visualized attributes into account, the musicologist is now able to measure possible pathways, especially by observing shared places of activity (see Figure 5). Although the musicologist requires additional literature to examine this question more precisely, the system provides valuable evidence to narrow the number of possibilities.

7 Discussion

The proposed profiling system was designed to support answering a novel type of research question in musicology. Some aspects of the collaborative work are outlined below.

Evaluation When developing the profiling system, we closely collaborated with four musicologists – a professor, a PhD student and two M.Sc. students –, who iteratively evaluated current prototypes. One of the key features of the profiling system was the design of similarity measures for relevant musician attributes included in the profiling process. Some of the similarity measures were refined step-by-step to incorporate musicological knowledge in order to gain results that meet the expectations of the musicologists. For example, when designing the activity region similarity, we always provided a list of place tuples with their calculated similarities to the musicologists to ensure an appropriate representation of musicological imagination of space. When determining relationship similarity – first defined for two musicians only by their distance in the social network graph – we mapped relationship strengths, provided by the musicologists, to edge lengths. As a result, familial relationships form clusters, which was an important requirement of the musicologists. Activity time similarity – first defined as lifetime similarity by birth and death of musicians – was also iteratively modified. Here, the inclusion of *first mentioned* dates and the mapping of uncertainties allowed us to define this similarity measure more precisely. The visualization of the profiling system was also iteratively improved and evaluated by the musicologists to meet their needs. This included both aspects of visual representation and interaction design. We could communicate our own concerns as well. For instance, we thought that overlapping streams in the vertical timeline are too confusing. But the musicologists prevented us from changing this representation arguing that it perfectly reflects their imagination of activity time similarity. As there is no ground truth regarding the profiling of musicians, the accuracy of our approach is not easy to measure. Sometimes, surprising and unexpected results occur. Being involved in all development stages, the musicologists assess individual similarity measures as well as entire profiling results as reasonable, which underpins the benefit of our method for musicology.

Limitations Being a challenge for developing the visual analytics system on the one hand, the existence of uncertain temporal metadata slightly affects the reliability of a profiling result. As the BMLO gets updated gradually, the removal of uncertainties requires future effort for musicologists using the database. A further limitation concerns the missing consideration of historical circumstances when calculating activity region similarities. First, the meaning of a geographical distance varies for different ages. Whereas a travel between European cities required several weeks in the Renaissance era, such a trip takes only few hours nowadays. Second, the usage of contemporary political conditions cannot be applied appropriately to historical contexts, although our collaborative solution turned out to be heuristically valuable for musicologists. But the elaboration of historical place identifiers could further improve the profiling result. As the provided textual metadata is not linked, e.g., the existence of “London” as activity place, “bandmaster” as musical profession and “church” as division does not imply that a musician indeed was a bandmaster at a church in London. The interpretation of such information still requires a musicologist’s knowledge or the usage of further sources. In terms of scalability, our proposed system is designed to compare the profiles of a rather small number of musicians – distinguishable through various colors –, usually less than ten. Therefore, general research interests like analyzing and comparing all *court composers* is not supported.

Future Work The BMLO is an ongoing digital humanities project under crowdsourcing aspects. Next to potential future data transformation and data representation challenges, the collaborating musicologists suggested several improvement prospects to determine similar musicians more precisely. First, the inclusion of hierarchical information into the profiling process was an often discussed issue. At the moment, a hierarchy is only given for musical professions. According to the musicologists, hierarchical representations for other text-based metadata dimensions (further professions, divisions), hierarchical relationship strengths or the calculation of activity region similarities taking historical circumstances into account could further strengthen the result of a profiling process. Consequently, we would need to adapt similarity measures and the visualization, especially the Column Explorer. Another future work is the profiling for coupled musicians – a novel type of research question stimulated through our profiling system. For instance, the profiling result for musicians similar to Wolfgang Amadeus Mozart and Johann Sebastian Bach in a single request could answer the question if a found musician would be more similar to Mozart or to Bach. As a straightforward adaption of the similarity measures does not anticipate adequate results, we require further interdisciplinary sessions to discuss required implementation steps.

8 Conclusion

As of the late 19th century, musicology focuses primarily on fifty musicians and their main works in a traditional philological manner. The achievements of other musicians only obtain less attention. The proposed profiling system aims to change this imbalance by rather throwing the spotlight on less popular musicians. Based upon a musician of interest – potentially one of the popular ones – musicologists are now capable of discovering less popular musicians with similar careers.

During the development, we closely collaborated with musicologists, who state that the resultant profiling system is a valuable analysis instrument that serves a novel type of research interest and provokes new research questions. Thereby, we designed the profiling system the way that it can easily be adapted to other historical groups of people.

Our presented approach facilitates comparative methods and research questions concerning musicians – for the first time with the aid of visual means. As the visualization indicates historical circumstances and cultural contexts, it gets possible to review time-dependent ideological opinions about individual musicians. Usage scenarios showcasing Handel’s, Spieß’ and Kohn’s careers demonstrate this capability of the profiling system.

References

[1] Bayerisches Musiker-Lexikon Online, 2015. ed. Josef Focht. `http://www.bmlo.lmu.de/` (Accessed 2015-03-31).

[2] A. Abdul-Rahman, J. Lein, K. Coles, E. Maguire, M. Meyer, M. Wynne, C. R. Johnson, A. Trefethen, and M. Chen. Rule-based Visual Mappings– with a Case Study on Poetry Visualization. In *Computer Graphics Forum*, volume 32, pages 381–390. Wiley Online Library, 2013.

[3] G. Andrienko, N. Andrienko, C. Hurter, S. Rinzivillo, and S. Wrobel. From Movement Tracks through Events to Places: Extracting and Characterizing Significant Places from Mobility Data. In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*, pages 161–170. IEEE, 2011.

[4] N. Andrienko and G. Andrienko. *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005.

[5] S. Bateman, C. Gutwin, and M. Nacenta. Seeing Things in the Clouds: The Effect of Visual Features on Tag Cloud Selections. In *Proceedings of the Nineteenth ACM Conference on Hypertext and Hypermedia*, HT ’08, pages 193–202. ACM, 2008.

[6] P. J. Besl and N. D. McKay. Method for Registration of 3-D Shapes. In *Robotics-DL tentative*, pages 586–606. International Society for Optics and Photonics, 1992.

[7] A. Bezerianos, P. Dragicevic, J. Fekete, J. Bae, and B. Watson. GeneaQuilts: A System for Exploring Large Genealogies. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1073–1081, Nov 2010.

[8] M. Brehmer, S. Ingram, J. Stray, and T. Munzner. Overview: The Design, Adoption, and Analysis of a Visual Document Mining Tool for Investigative Journalists. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12):2271–2280, Dec 2014.

[9] L. Byron and M. Wattenberg. Stacked Graphs – Geometry & Aesthetics. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1245–1252, Nov 2008.

[10] M. Cammarano, X. Dong, B. Chan, J. Klingner, J. Talbot, A. Halevy, and P. Hanrahan. Visualization of Heterogeneous Data. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1200–1207, Nov 2007.

[11] H.-C. Chen and A. L. Chen. A music recommendation system based on music data grouping and user interests. In *Proceedings of the tenth international conference on Information and knowledge management*, pages 231–238. ACM, 2001.

[12] J. Chen, W. Geyer, C. Dugan, M. Muller, and I. Guy. "Make New Friends, but Keep the Old" – Recommending People on Social Networking Sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 201–210. ACM, 2009.

[13] J. Choo, C. Lee, H. Kim, H. Lee, Z. Liu, R. Kannan, C. Stolper, J. Stasko, B. Drake, and H. Park. VisIRR: Visual Analytics for Information Retrieval and Recommendation with Large-Scale Document Data. In *Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on*, pages 243–244, Oct 2014.

[14] C. Collins, F. Viegas, and M. Wattenberg. Parallel Tag Clouds to Explore and Analyze Faceted Text Corpora. In *Visual Analytics Science and Technology, (VAST), 2009 IEEE Symposium on*, pages 91–98, Oct 2009.

[15] M. Correll, M. Witmore, and M. Gleicher. Exploring Collections of Tagged Text for Literary Scholarship. *Computer Graphics Forum*, 30(3):731–740, 2011.

[16] W. Cui, S. Liu, Z. Wu, and H. Wei. How Hierarchical Topics Evolve in Large Text Corpora. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12):2281–2290, Dec 2014.

[17] J. Davidson, B. Liebald, J. Liu, P. Nandy, T. Van Vleet, U. Gargi, S. Gupta, Y. He, M. Lambert, B. Livingston, et al. The YouTube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 293–296. ACM, 2010.

[18] S. Debnath, N. Ganguly, and P. Mitra. Feature Weighting in Content Based Recommendation System Using Social Network Analysis. In *Proceedings of the 17th international conference on World Wide Web*, pages 1041–1042. ACM, 2008.

[19] E. W. Dijkstra. A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik*, 1(1):269–271, 1959.

[20] M. Dörk, S. Carpendale, C. Collins, and C. Williamson. VisGets: Coordinated Visualizations for Web-based Information Exploration and Discovery. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1205–1212, Nov 2008.

[21] M. Dörk, N. Riche, G. Ramos, and S. Dumais. PivotPaths: Strolling through Faceted Information Spaces. *Visualization and Computer Graphics, IEEE Transactions on*, 18(12):2709–2718, Dec 2012.

[22] W. Dou, X. Wang, D. Skau, W. Ribarsky, and M. Zhou. LeadLine: Interactive Visual Analysis of Text Data through Event Identification and Exploration. In *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on*, pages 93–102, Oct 2012.

[23] E. Gansner, Y. Hu, S. Kobourov, and C. Volinsky. Putting Recommendations on the Map: Visualizing Clusters and Relations. In *Proceedings of the Third ACM Conference on Recommender Systems*, RecSys '09, pages 345–348, New York, NY, USA, 2009. ACM.

[24] Z. Geng, T. Cheesman, R. S. Laramee, K. Flanagan, and S. Thiel. ShakerVis: Visual analysis of segment variation of German translations of Shakespeare's Othello. *Information Visualization*, page 1473871613495845, 2013.

[25] F. Gibbs and T. Owens. Building Better Digital Humanities Tools: Toward broader audiences and user-centered designs. *Digital Humanities Quarterly*, 6(2), 2012.

[26] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry. Using Collaborative Filtering to Weave an Information Tapestry. *Communications of the ACM*, 35(12):61–70, 1992.

[27] B. Gretarsson, J. O'Donovan, S. Bostandjiev, C. Hall, and T. Höllererk. SmallWorlds: Visualizing Social Recommendations. In *Proceedings of the 12th Eurographics / IEEE - VGTC Conference on Visualization*, EuroVis'10, pages 833–842, Aire-la-Ville, Switzerland, Switzerland, 2010. Eurographics Association.

[28] I. Guy, N. Zwerdling, I. Ronen, D. Carmel, and E. Uziel. Social Media Recommendation Based on People and Tags. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '10, pages 194–201, New York, NY, USA, 2010. ACM.

[29] M. Harrower and C. A. Brewer. ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. *The Cartographic Journal*, 40(1):27–37, 2003.

[30] K. Head. Gravity for Beginners. *University of British Columbia*, 2003.

[31] J. Heer and D. Boyd. Vizster: Visualizing Online Social Networks. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pages 32–39, Oct 2005.

[32] F. Heimerl, S. Koch, H. Bosch, and T. Ertl. Visual Classifier Training for Text Document Retrieval. *Visualization and Computer Graphics, IEEE Transactions on*, 18(12):2839–2848, Dec 2012.

[33] Y. Jang, A. Malik, D. S. Ebert, R. Maciejewski, W. Huang, and N. Elmqvist. A Correlative Analysis Process in a Visual Analytics Environment. In *Proceedings of the 2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*, VAST '12, pages 33–42, Washington, DC, USA, 2012. IEEE Computer Society.

[34] S. Jänicke, G. Franzini, M. F. Cheema, and G. Scheuermann. On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges. In R. Borgo, F. Ganovelli, and I. Viola, editors, *Eurographics Conference on Visualization (EuroVis) - STARs*. The Eurographics Association, 2015.

[35] S. Jänicke, A. Geßner, M. Büchler, and G. Scheuermann. Visualizations for Text Re-use. *GRAPP/IVAPP*, pages 59–70, 2014.

[36] S. Jänicke, C. Heine, R. Stockmann, and G. Scheuermann. Comparative Visualization of Geospatial-temporal Data. In *GRAPP/IVAPP*, pages 613–625, 2012.

[37] D. Keim and D. Oelke. Literature Fingerprinting: A New Method for Visual Literary Analysis. In *Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on*, pages 115–122, Oct 2007.

[38] S. Koch, M. John, M. Worner, A. Muller, and T. Ertl. VarifocalReader – In-Depth Visual Analysis of Large Text Documents. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12):1723–1732, Dec 2014.

[39] J. A. Konstan and J. Riedl. Recommender systems: from algorithms to user experience. *User Modeling and User-Adapted Interaction*, 22(1-2):101–123, 2012.

[40] L. Lü, M. Medo, C. H. Yeung, Y.-C. Zhang, Z.-K. Zhang, and T. Zhou. Recommender Systems. *Physics Reports*, 519(1):1–49, 2012.

[41] T. Munzner. A Nested Model for Visualization Design and Validation. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6):921–928, 2009.

[42] M. J. Pazzani and D. Billsus. Content-Based Recommendation Systems. In *The adaptive web*, pages 325–341. Springer, 2007.

[43] A. Perer and B. Shneiderman. Balancing Systematic and Flexible Exploration of Social Networks. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):693–700, Sept. 2006.

[44] N. Riche and T. Dwyer. Untangling Euler Diagrams. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1090–1099, Nov 2010.

[45] K. Sachs-Hombach. Das Bild als kommunikatives Medium. *Elemente einer allgemeinen Bildwissenschaft*, 1993.

[46] L. Shi, F. Wei, S. Liu, L. Tan, X. Lian, and M. Zhou. Understanding Text Corpora with Multiple Facets. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, pages 99–106, Oct 2010.

[47] Y. Shrinivasan, D. Gotz, and J. Lu. Connecting the Dots in Visual Analysis. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on*, pages 123–130, Oct 2009.

[48] J. Stasko, C. Gorg, Z. Liu, and K. Singhal. Jigsaw: Supporting Investigative Analysis through Interactive Visualization. In *Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on*, pages 131–138, Oct 2007.

[49] K. Verbert, D. Parra, P. Brusilovsky, and E. Duval. Visualizing Recommendations to Support Exploration, Transparency and Controllability. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces*, IUI '13, pages 351–362, New York, NY, USA, 2013. ACM.

[50] M. Vlachos and D. Svonava. Recommendation and visualization of similar movies using minimum spanning dendrograms. *Information Visualization*, page 1473871612439644, 2012.

[51] R. Vuillemot, T. Clement, C. Plaisant, and A. Kumar. What's being said near "Martha"? Exploring name entities in literary text collections. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on*, pages 107–114, Oct 2009.

[52] M. Wattenberg and F. Viegas. The Word Tree, an Interactive Visual Concordance. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1221–1228, Nov 2008.

[53] C. Weaver. Multidimensional Visual Analysis Using Cross-Filtered Views. In *Visual Analytics Science and Technology, 2008. VAST '08. IEEE Symposium on*, pages 163–170, Oct 2008.

[54] C. Weaver. Multidimensional Data Dissection Using Attribute Relationship Graphs. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, pages 75–82, Oct 2010.