International Journal of Art, Culture and Design Technologies
Volume 8 • Issue 1 • January-June 2019

Data-Driven Maps of Art History

Doron Goldfarb, Vienna University of Technology, Vienna, Austria
Dieter Merkl, Vienna University of Technology, Vienna, Austria

ABSTRACT

Many approaches to explain conceptions of developments in the arts to both peers and a more general audience included diagrams representing networks of related entities such as genealogies of important historical actors or styles. While such visualizations were traditionally created by hand, the recent emergence of extensive digital repositories of art history information enable new means of presentation. This work seeks to explore the potential of openly available data sources to create bottom-up, data driven versions of such network maps of art history. It highlights commonalities and differences between views derived from institutional and crowd-sourced data repositories and compares them with identified historical examples. The results suggest that the available data can be used to create largescale views on specific developments in art history, potentially serving as aid for navigating vast online collections of digitized artworks but also as means of reflection on the origin of the data sources themselves.

KEYWORDS

Art Styles, Biographies, Core-Periphery, Data Analytics, DBpedia, Genealogy, Network Visualization, ULAN, Wikidata, Wikipedia

INTRODUCTION

Especially after the inception of art history as an academic discipline in the late 18th century, diagrams have been used as visual devices to convey views on related developments. They were included in publications and textbooks and served as communicational means to support scholarly positions and as educational tools providing bird's eye views on large-scale historical developments. Famous examples such as the "Diagram of Stylistic Evolution from 1890 until 1935" from (Barr, 1986/1937) sought to present visual aggregations of multitudes of individual processes as macroscopic, law-like views on artistic evolution. As discussed in (Schmidt-Burkhardt, 2005), such charts were often praised for their educational value but also challenged for lacking objectivity due to omissions or special foci introduced by their authors.

In recent years, the increasing availability of both digitized and born-digital art history resources went in parallel with the emergence of new paradigms of data-driven analysis which can be subsumed under the umbrella term "data analytics". The emerging field of digital art history increasingly uses algorithmic methods to analyze the accumulating body of historical material, while recent trends in museology propose digital means of presenting art collections to an increasingly tech-savvy audience,

DOI: 10.4018/IJACDT.2019010101

seeking to contextualize exhibits using multimedia guides and interactive installations for local and dedicated online presentations for remote visitors. Information visualization plays an important role in both scenarios, supporting researchers in the interpretation of large data collections and the communication of their findings, but also providing virtual visitors with navigational guides for finding their way through vast online art collections.

The original aim of this work was to explore the potential of historical social network data, represented through interlinked artist biographies, for providing bird's eye overviews on art historical developments beyond the relatively limited scope of existing, often manually created maps of art history. This was driven by the expectation that large-enough amounts of recorded ties between artists and/or other important persons from the art world would aggregate into large-scale networks spanning multiple centuries whose visualization could support the contextualization of artworks in virtual presentations. The analysis of existing biographical data sources, however, revealed that they were not free of cultural/institutional biases which appeared in form of differing compositions of nationalities and/or roles of covered persons. This provided an additional and relevant research perspective which directly linked to existing discussions about modes of inclusion and exclusion in the formation of the canon of art history, global aspects of which were often subsumed under the notion of the core and the periphery of the art world (Joyeux-Prunel, 2014).

The aims of the work presented in this article were therefore (1) to demonstrate the possibility of creating data-driven maps of art history and (2) to show their use for revealing specific biases present in different data collections. This paper unifies the authors' previous related work (Goldfarb, Arends, Froschauer, & Merkl, 2013) (Goldfarb, Arends, Froschauer, Weingartner, & Merkl, 2014) under a common data processing framework and discusses large-scale network visualizations of art history biographies generated with content from the Getty Union Artist Names and Wikipedia, comparing them with each other and with existing scholarly examples. Different filtering approaches are used to highlight data specific aspects, including means to unravel chronological structure embedded in highly interlinked sets of historical entities and to reveal hidden interactions between subgroups of them. The results show that data-driven maps of art history successfully extend their manually created scholarly counterparts by putting them in larger historical context and at the same time serve as tool to reflect upon the nature of the used data sources themselves.

DATA SOURCES AND TOOLS

The Getty Union List of Artist Names (ULAN) is an established data source curated by domain experts as controlled vocabulary for person names relevant for art history. As described in (Baca & Gill, 2015), it has recently been made available as Linked Data, reflecting the current trend for cultural institutions to open their databases for public access and re-use. Besides containing records for about 200,000 persons and institutions, the former with basic information such as nationality, role, birth and death dates, etc., the ULAN also provides associative links between a subset of its records, representing identified professional, familial and other relationship types. Due to its institutional background, the presence of persons in the ULAN can be considered as evidence for their importance for the domain of art history.

As one of the premier examples for freely accessible, crowd-curated content on the Internet, Wikipedia in turn unites a wide range of contributions from both domain experts and laypersons, mainly relying on mutual reviews for ensuring the quality of provided content. Being an encyclopedia, its articles are usually dedicated to distinct entities such as persons, places or concepts and are mutually connected via hyperlinks. This allows to treat interlinked biographies on Wikipedia like a social network and to compare it with more explicit network structures such as found in the ULAN. Several projects and initiatives make Wikipedia data accessible for quantitative use: Preprocessed and yearly updated page-link datasets extracted from various Wikipedia language versions are provided by DBpedia, an approach to extract structured information from Wiki pages (Auer et al., 2007). The

Wikidata platform in turn features structured records with community contributed facts about entities present in Wikipedia and includes many person records interlinked with their respective Wikipedia biographies in multiple languages (Vrandečić & Krötzsch, 2014). Due to continuous efforts of the Wikipedia community to attach authority control information to Wikipedia articles, described in (Klein & Kyrios, 2013), many of these person records contain ULAN mappings which can be used as reference to identify Wikipedia biographies relevant to art history.

In this work, Wikidata, DBpedia and the ULAN were used to derive three related datasets, including (1) the network of ULAN biographies connected via associative ULAN links, (2) the network of multilingual Wikipedia biographies mapped to the ULAN via a dedicated Wikidata property (*ID: P245*) and interlinked via multilingual Wikipedia hyperlinks derived from DBpedia and (3) the bipartite network of DBpedia links between multilingual articles about art/architecture styles and Wikipedia biographies, all identified via Wikidata entities tagged as art movement (*ID: Q968159*), architectural style (*ID: Q32880*) or human (*ID: Q5*), the latter not limited to those mapped to the ULAN. Datasets (1) and (2) were used for the person network experiment described below, followed by the person-style network experiment using dataset (3).

All the presented network visualizations are based on the ForceAtlas 2 algorithm described in (Jacomy, Venturini, Heymann, & Bastian, 2014) as implemented in the Gephi platform. This layout algorithm turned out to be most suited for visualizing networks with intrinsic chronological features. The R statistics platform was used for data processing and analysis.

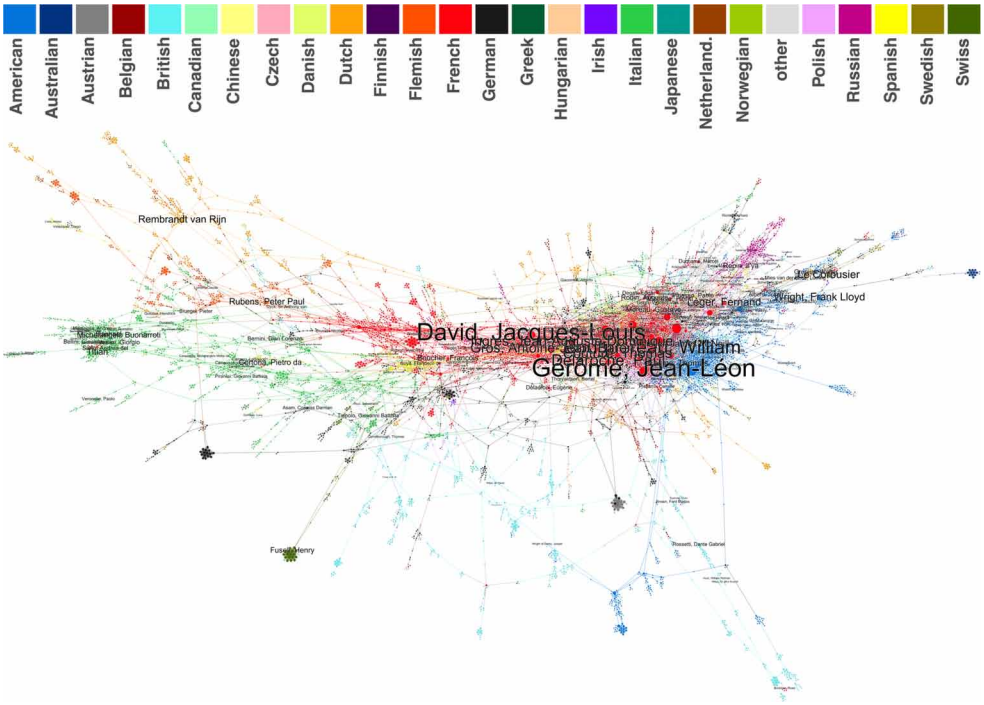
PERSON NETWORKS IN THE ULAN AND IN WIKIPEDIA

ULAN Person Network

The ULAN dataset of summer 2015 connected 21,942 person records via 50,076 unique but mirrored – e.g. all student_of links had complementary teacher_of links – associative links, corresponding to 25,038 unique pairs of interlinked ULAN records. Having 4,083 connected components (CC) in total, the network featured a giant connected component (GCC) consisting of 10,444 (47.58%) persons interlinked via 29,956 (59.82%) links. Since the next largest component only featured 39 persons, the GCC was the focus of the analysis and its visualization is shown in Figure 1. The persons featured in the chart were found to follow a rough chronological succession from left to right, with Renaissance and Baroque artists dominating the left side, late 19th and 20th century persons the right. The color alphabet proposed in (Green-Armytage, 2010) was used for color-coding the person nodes by the top 26 (plus one for "other") preferred nationalities assigned in the ULAN. This even more underscored the chronological nature of the network by revealing clear-cut national clusters of persons, especially visible for the three largest nationalities representing about 50.5% of the persons in the GCC — Italian (14.33%), French (22.71%) and American (13.47%). Their apparent chronological succession reflected shifting cultural centers of the art world through time. Additional developments appeared above and below this main strand, such chains of loosely connected groups of Dutch (6.26%) and Flemish (3.16%) persons on top, many of them from the Dutch Golden Age, and a similarly loose group of British (8.56%) persons at the bottom. Another large group of persons, Germans (8.35%), was spread across the whole plot without forming a contiguous section. Many densely connected, clique-like subgroups of persons were found to represent families.

The densely documented relationships for Italian, French and American persons stood in contrast to the rather "skinny" networks recorded for other nationalities, which suggested that these groups of people were the focus of the collections present in the institutions contributing to the ULAN, an issue to be taken into account when considering these data for creating a large-scale view on art historical developments over centuries. Fundamental art historical epochs such as Italian Renaissance, Dutch Golden Age, French Academism, etc., nevertheless became clearly visible in the visualization of the ULAN network, which was found to be in interesting agreement with the "Chart of Art History" produced by Newton in 1941 for the textbook "European Painting and Sculpture" (Newton, 1956/1941,

Figure 1. Visualization of the ULAN network by nationality, colored by nationality



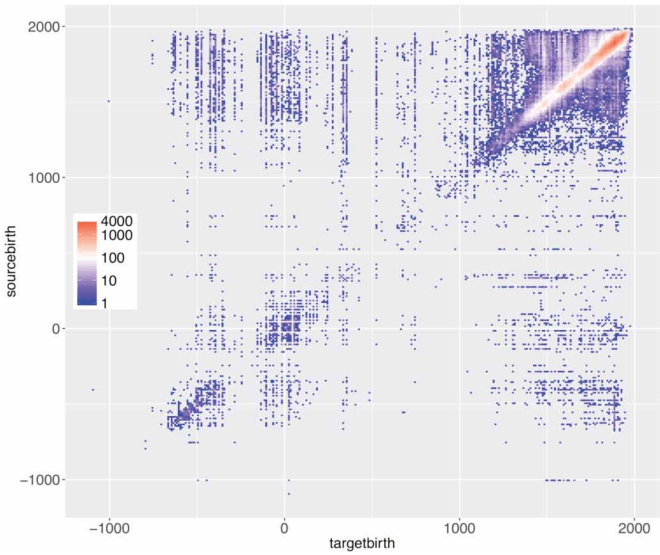
p. 237). Quite like the ULAN network, the manually created chart featured the centuries between 1255 to 1900 to be dominated by Italian and French artists which outnumbered the representatives of all the other featured European schools. Overall, the visualization of the ULAN GCC showed that biographical network data could in principle be used to provide large-scale views on historical developments and the ULAN's apparent focus on specific nationalities provided a strong incentive to explore additional sources, such as Wikipedia, for comparison.

Person Networks in Wikipedia

As of March 2016, 47,463 out of 55,122 Wikidata person records with ULAN mappings had a corresponding Wikipedia article in at least one language and 40,943 of them were connected via 283,368 unique Wikipedia hyperlinks as identified in one or more of the 50 largest DBpedia language versions. In contrast to the ULAN, Wikipedia hyperlinks between person biographies were not always mirrored, the identified 283,368 hyperlinks represented 237,139 unique pairs of ULAN mapped biographies. Although still very sparse, the multilingual network of Wikipedia hyperlinks thus had higher density than the ULAN associative link network (5.79 vs. 1.14 average links per person, Link density 0.000283 vs 0.000104) and its GCC contained 99.08% of the 40,943 involved persons. 12,239 (55,78%) of the ULAN person records interconnected via associative links in the ULAN were found to have corresponding Wikipedia articles mutually interlinked in at least one of the 50 observed Wikipedia language versions and 80.72% of the existing ULAN associative links between them were also represented as Wikipedia hyperlinks.

The overlap between the ULAN and the Wikipedia network suggested that an intrinsic chronology should also exist in the latter, but an initial visualization exposed no visible chronology there, prompting for a more thorough exploration. In order to analyze the temporal distribution of Wikipedia hyperlinks between the mapped ULAN persons, the distributions of the source and target birth dates

Figure 2. 2D histogram of birth dates of interlinked Wikipedia biographies

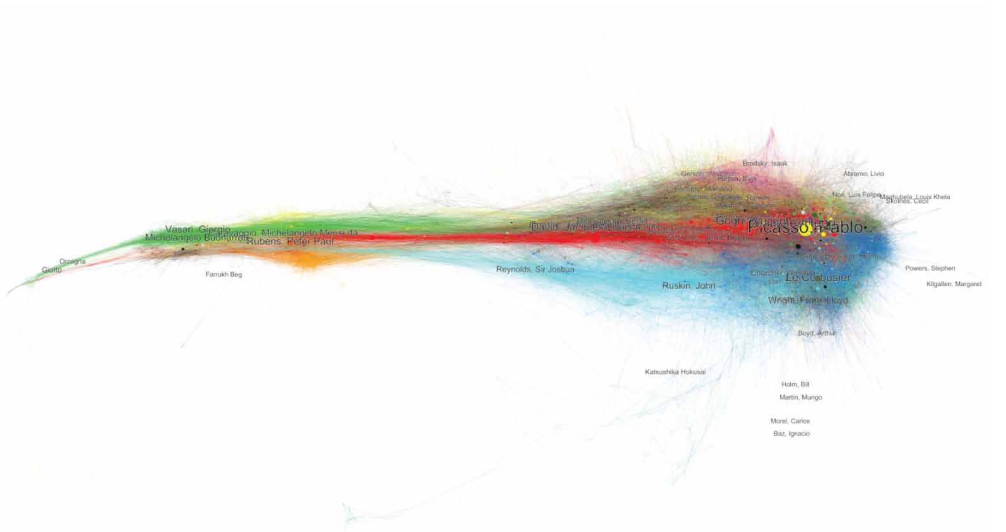


of the linked pairs were analyzed and plotted on a 2D histogram shown in Figure 2, revealing various temporal characteristics of the structure of the Wikipedia hyperlink network. Generally increasing in numbers across time and in parallel with the overall distribution of persons birth dates peaking at around 1850, about 85% of all the hyperlinks connected persons born not more than 75 years apart. Moreover, about two thirds of the Wikipedia hyperlinks pointed from biographies of younger persons to those of older persons. One reason for this temporal asymmetry were articles about outstanding artists and other important persons which were far more often referenced by biographies of later generations than vice versa. Some biographies, e.g. articles about important teachers, nevertheless featured relatively long lists of links to their pupils, which in turn contributed to the still extensive number of hyperlinks pointing into the future. Finally, there was an apparent discontinuity between groups of persons from Greek and Roman antiquity, visible as two distinct blobs at the diagonal centered around the years 500 BC and 0 AD, and the ”main” group of persons from the late middle ages until today. Separated by a clearly visible gap during the middle ages, their only connection via links of long temporal distance reflected the renewed interest in the culture and the arts of antiquity at the onset of the modern age.

Concerning the identified temporal characteristics, the main difference between the ULAN and Wikipedia networks were that the former consisted of symmetric links mainly between contemporaries (95.55% born within 75, 98% within 100 years of each other), which resulted in persons from antiquity to be disconnected from the ULAN GCC. This suggested that using a cutoff of 75 years to remove “long duration” links from the Wikipedia network could uncover the temporal succession of the featured biographies via the force-based layout, although resulting in a similar “lost connection” to antiquity. This was confirmed in Figure 3 which shows the visualization of the GCC of the filtered network combined from all 50 DBpedia language versions, featuring only persons born after 1000 AD. Using the same color-coding for nations as used in Figure 1 for the ULAN GCC revealed a similar chronological succession of distinct art historical periods.

Since the selection of Wikipedia biographies was based on persons featured in the ULAN, it was not surprising to find a similar succession of Italian (green), French (red) and American (blue) clusters in the filtered Wikipedia network, which nevertheless also contained dense clusters of other nationalities. Many Dutch/Flemish persons (light/dark orange) for example were concentrated in a

Figure 3. Pruned network of Wikipedia biographies identified via the ULAN in 50 language versions, colored by nationality

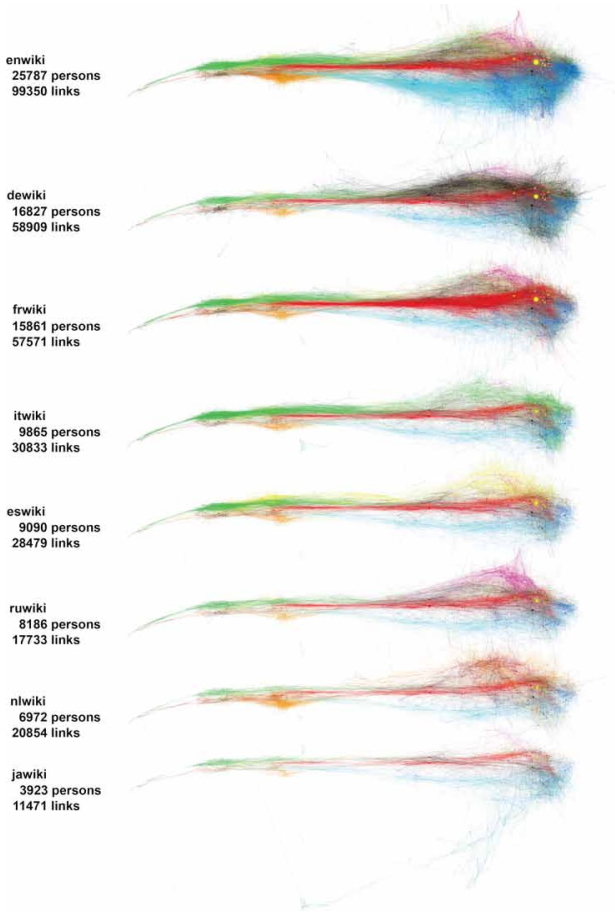


"Dutch Golden Ag" cluster, while British persons (light blue) appeared in a parallel development to the central French "strand". Clusters of other nationalities in turn overlapped in the upper part of the Figure to a high degree.

While the visualization of the combined Wikipedia biography network showed a relatively balanced presence of the different nationalities, most of the 50 individual contributing language versions had a clear tendency to focus on culturally related content. In order to highlight this observation, Figure 4 shows eight individual networks representing the seven largest contributing language versions (English, German, French, Italian, Spanish, Russian and Dutch) and the Japanese version as largest non-European language, all created using the same layout as used in Figure 3 for better comparison. The separated views emphasized that each Wikipedia version featured especially dense person clusters for those nationalities related to its language, which was often due to unique person biographies not present elsewhere and sometimes revealed remarkable developments. The distinct Japanese strand (turquoise) shown at the bottom of the Japanese Wikipedia network visualization for example had only minimal connections to other cultures until the late 19th century Meiji era, whose opening towards the West became dramatically visible in the plot.

Regardless of the emphasis put on their own cultures, the observed language versions also featured many similarities. Most of them shared biographies about the most famous representatives of art history whose mutual ties formed a recurring "skeleton" found in all the different language versions and also in the ULAN associative link network, spanning from the Renaissance (Giotto, Da Vinci, Raphael, Dürer, ..) and the Baroque (Caravaggio, Rembrandt, Rubens, ..) across the 18th and 19th centuries (Goya, David, Delacroix, Courbet, Monet, ..) to modern (Cézanne, Matisse, Kandinsky, Picasso, Dalí, ..), post war (Pollock, Warhol, ..) and contemporary art (Koons, Hirst, Banksy, ..). Interesting visual analogies were found in two different historical diagrams: Similar to the visualizations of the national clusters in the filtered Wikipedia networks, a concurrent presence of interconnected cultural strands was found in (Strass, 1828/1804), a remarkable manually created visualization of the interplay between different world cultures across history as parallel, merging and bifurcating rivers. Alfred H. Barr's curatorial "torpedo" diagrams from (Barr, 1941), sketching the composition of the ideal MoMA collection, represented another striking analogy: The filtered Wikipedia biography network visualizations showed similar snapshots of the interplay of changing centers of cultural hegemony

Figure 4. Pruned biography networks as present in eight individual Wikipedia languages, colored by nationality



across time, but this time not as determined by a single art historian but via the collection of ULAN biographies filtered by language-specific coverages of different Wikipedia versions.

Network of Aggregated Links Between Nationalities

The visualizations of the network of ULAN biographies in Wikipedia highlighted the chronological interactions between groups of persons of different nationalities. The high number of homogeneous relationships between persons belonging to one of the few top counting nationalities (the "core"), however, often masked the interactions between persons of less prominent origin (the "periphery"). One attempt to highlight interactions in or with the periphery was to move from the individual-based micro-level to a macro-level network of interacting nationalities, derived from aggregating nation-to-nation interactions from the individual level. This macro-level network consisted of nodes representing nationalities, all pairs of which connected via links weighted by the absolute counts of individual-level links between persons of the respective nationalities. The observed counts, however, were still highly skewed towards the core nations and interactions in the periphery thus assigned very low overall weights, calling for a relative measure for weighting the aggregated nation-to-nation ties taking the size of each nationality into account.

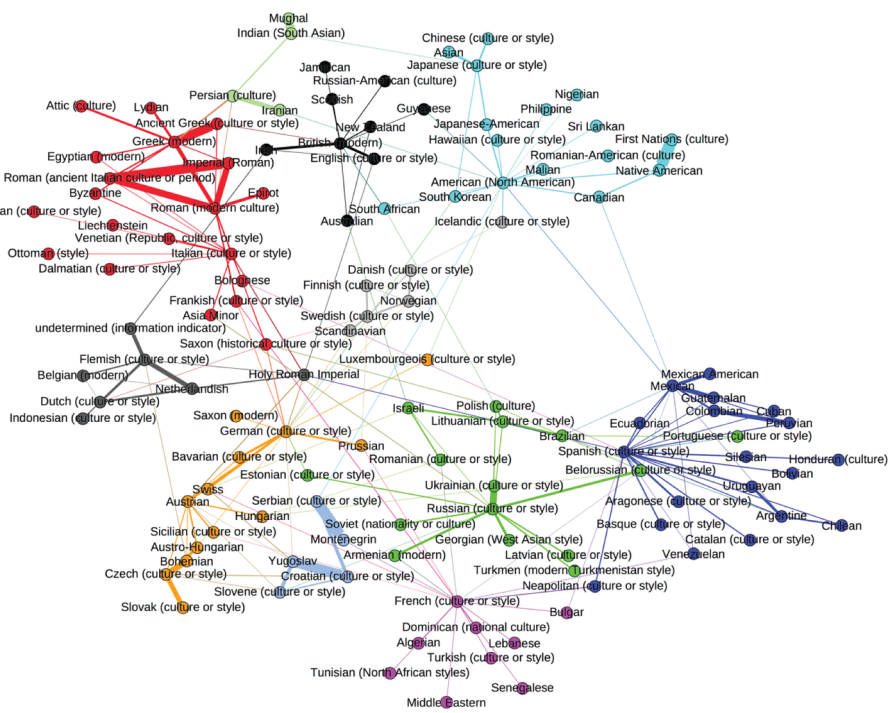
An existing approach could be identified for this purpose, based on "standardized residuals" as described in (Agresti, 2007, p. 38) in the context of Pearson's Chi-Squared test for the independence

of two categorical variables. The 34,710 possible pairs of 195 source and 178 target nationalities of the ULAN mapped and interconnected Wikipedia biographies were thus arranged in a contingency table holding the hyperlink counts between each nation pair as observed in the combined Wikipedia language network. In order to reduce effects from noise, counts below five were set to zero before calculating the standardized residuals as described in (Agresti, 2007). The results were filtered for a minimum of three positive standardized deviations from expectation, yielding a network of 371 interactions between 128 distinct nationalities. Figure 5 shows a visualization of this network where the thickness of each tie represents the value of its standardized residual, now revealing interesting connections in and with the periphery. There was an apparent, clustered structure of different nationality groups each having highly interconnected members.

The clustering could be confirmed by applying the Louvain community detection method described in (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) which identified 14 highly separated groups with a high modularity value, as defined in (M. E. J. Newman, 2006), of 0.736. The identified groups are shown in different colors in Figure 5 and it was interesting to observe that in many cases, they were based on former colonial ties, especially visible for the French (bottom center, purple), Spanish (lower right, blue) and British (top center, black) spheres of influence. Other groups were based on changing state organizations throughout history, such as the group of former Yugoslavian countries (lower center, light blue), the Russian/Soviet sphere of influence (lower center, green) or the Austrian/German group including former members of the Austro- Hungarian empire (lower left center, orange). Some of the clusters also linked ancient and middle age "nationalities" to their modern counterparts, such as those containing Italians and Greeks (top left, red) or Iranians and Indians (top center, light green).

Although clearly distinct from each other, the different encountered clusters revealed varying levels of homogeneity. While the cluster for Scandinavian nations (center, grey) clearly reflected the tight relationships between the involved countries, the cluster of Asian and North American (top right, cyan)

Figure 5. Residual network of aggregated nation to nation links



was less homogeneous, suggesting that the community detection algorithm was not able to detect all the clusters potentially present in the observed dataset. Overall, the clusters of nationalities revealed interesting associations with a network of Wikipedia languages studied in (Samoilenko, Karimi, Edler, Kunegis, & Strohmaier, 2016). The authors found cultural clusters of Wikipedia language editions and used contemporary data sources about country features such as religion and spoken languages to explain their formation to be mainly based on bilinguality and shared religion. Although based on different types of entities, the clustering of contemporary and historical "nations" observed in this work appeared to be based on relatively similar socio-cultural factors, raising interesting questions for future work. Considering a comparable quantitative analysis for historical entities that existed only for certain durations and significantly changed their geographical and cultural extent over time, however, would require significant effort for acquiring or even creating usable data sources in this regard.

BIPARTITE NETWORK OF STYLES AND PERSONS LINKED IN WIKIPEDIA

The bottom up views on interactions between persons of different cultural origin derived from the ULAN and Wikipedia provided interesting insight into the high-level structure of developments in art history and showed analogies to historical scholarly diagrams. Many of the latter, such as the initially mentioned one from (Barr, 1986/1937), however, mainly connected different styles with each other and did not feature many individuals. Such formalist views on stylistic influences have been driven by positions such as the "art history without names", assuming that style exists beyond the individual genius and is representative for a specific culture at a specific time. Computational approaches to such conceptions of stylistic influences would call for advanced image analysis methods allowing to directly extract patterns of style and their mutual interactions from the artworks themselves. While such methods will certainly influence art history research in future, it is nevertheless of interest to analyze the existing historiography about art styles and periods for its high-level structure and explore bird's eye visualizations of developments represented there.

The identified characteristics of the large-scale, multi-lingual Wikipedia biography network thus motivated to combine it with additional information about art and architecture periods and styles, which was realized as a bipartite network consisting only of links between biographies and multi-lingual Wikipedia articles referenced from 426 distinct Wikidata entries such as "Renaissance" or "Cubism". The underlying assumption was that a large enough dataset of such links would yield a contiguous structure of styles interlinked via biographies, featuring a chronological succession as found for the biographical networks alone. The extraction of Wikipedia hyperlinks between the 426 identified style articles and any Wikipedia person biography from the 50 largest DBpedia language versions resulted in a combined dataset of 145,079 hyperlinks connecting 406 of the identified styles with 69,908 persons. Only about a quarter of the latter were amongst the 40,000+ ULAN biographies mapped to Wikipedia, suggesting the style-based identification to be an alternative way to find art-historically relevant persons in the encyclopedia.

Since a first visualization of the bipartite network did not show any apparent chronology, a temporal filtering approach like the one used for the person networks was considered. As it turned out, especially the temporal distributions of persons linked to broad stylistic periods such as Baroque or Renaissance often covered a very wide range across centuries and were not limited to only the persons alive during their actual duration. This suggested to filter the data by removing, for each style, the links from persons who were not alive during its actual historical period, which was complicated by missing beginning and ending dates in many of the Wikidata records about styles and required to derive the durations via the temporal distribution of interlinked persons instead. A simple approach was to extract "peaks" in the temporal distributions of persons linked to each style and to choose the area around the largest peak as candidate for its actual duration. Unfortunately, this was very imprecise and failed especially for "older" periods which were more often referenced by recent biographies than by those from their own time, calling for an alternative approach taking the generally skewed temporal

distribution of persons into account. This suggested to create a contingency table dividing style-person link counts into rows by style and columns by half-centuries representing the rounded average of the birth and death years (mid-life-year) of interlinked persons and to calculate standardized residuals as described in (Agresti, 2007). For each style, the half-century with highest positive residual and its neighborhood above a certain threshold were determined to be the style's duration.

Table 1 illustrates this procedure for a dummy example. Four styles have different temporal distributions of persons interlinked with them in Wikipedia. For each style, the leftmost value in each cell represents the observed count of interlinked persons having their "mid-life-year" in the respective half-century. The center value in each cell represents the expected count as derived from the marginal distributions (shown at the right and at the bottom of the Table), while the right value represents the resulting standardized residual derived from the deviation between observed and expected counts. Assuming a threshold of zero, the residuals marked with an asterisk show the peaking half-centuries and their above-threshold neighbors determining the extracted duration of the respective style. The second row shows an example for a style having more references from later born persons than contemporaries. The residual-based approach determines the lower counting contemporaries to be deviating more significantly from expectation and thus favors the "earlier" peak to represent the style's duration.

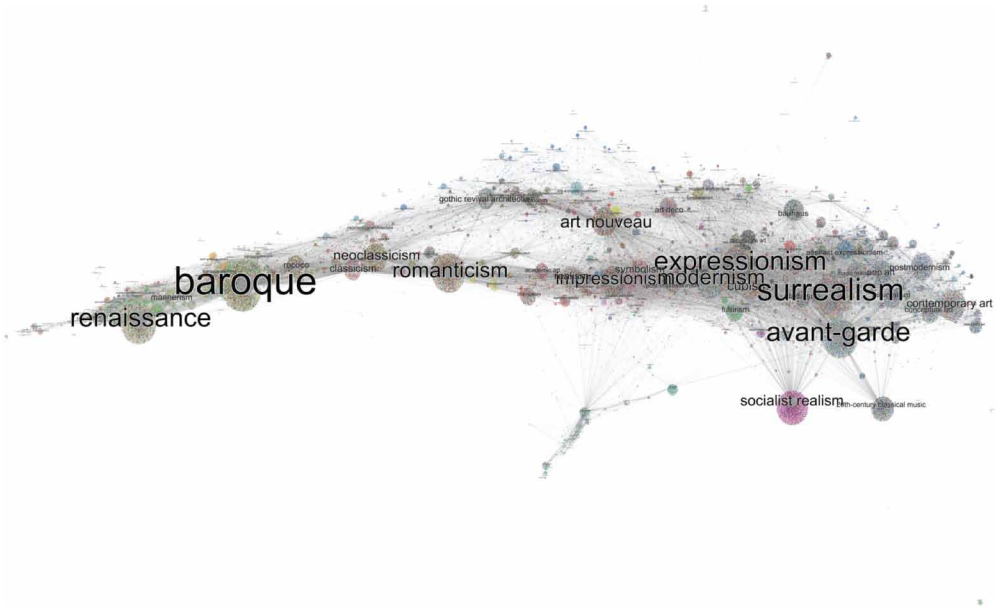
After removing 31 styles whose temporal distribution never exceeded two person links per half-century, the residual-based approach was evaluated against the count based one using manual annotations with beginning and end dates for the remaining set of 375 styles. Links both falling within the annotated style duration and the determined style duration were counted as "true positives", links only falling into the former as "false negatives", those only into the latter as "false positives", those outside both ranges as "true negatives". Both F1 as well as MCC measures ranked the residual-based approach to yield better combinations of precision and recall and suggested a quite low threshold of 0.33 for determining standardized residual peaks and their neighborhoods. For the 375 style periods, the manual annotations and the extracted beginning and end dates at least overlapped for 353 of them, while 22 were mismatched.

The extracted style durations were used to filter the network to a reduced set of 93,197 hyperlinks connecting the 375 styles with 55,499 persons and as shown in Figure 6, the pruned network now unfolded into an approximately chronological structure. A succession of styles became clearly visible, starting with (Pre-)Romanesque art and architecture, continuing with Gothic, Renaissance, Mannerist, Baroque styles until about the era of (Neo)Classicism, where the succession divided into relatively distinct strands, the top part mainly representing architectural, the lower part fine art styles. These two strands were nevertheless continuously interlinked via styles such as Art Nouveau, Art Deco and Bauhaus, which appeared as bridges between the now separated disciplines. The region beneath Bauhaus also contained the other modern art styles featured in Barr's "Diagram of Stylistic Evolution from 1890 until 1935" from (Barr, 1986/1937), showing that the data-driven version derived from Wikipedia successfully put them in a larger art historical context and also included styles which were

Table 1. Determining stylistic periods via standardized residuals

	1700	1750	1800	1850	1900	1950	
Style 1	0 / 3.46 / -2.35	0 / 5.35 / -3	5 / 3.46 / *1.04	10 / 5.35 / *2.61	20 / 12.91 / *2.9	5 / 9.45 / -2	40
Style 2	10 / 3.38 / *4.53	7 / 5.22 / *1.01	5 / 3.38 / *1.11	0 / 5.22 / -2.95	12 / 12.59 / -0.24	5 / 9.21 / -1.91	39
Style 3	1 / 1.91 / -0.75	10 / 2.94 / *4.86	0 / 1.91 / -1.59	5 / 2.94 / 1.42	6 / 7.1 / -0.55	0 / 5.2 / -2.87	22
Style 4	0 / 2.25 / -1.76	0 / 3.48 / -2.25	1 / 2.25 / -0.98	2 / 3.48 / -0.96	3 / 8.39 / -2.54	20 / 6.14 / *7.17	26
	11	17	11	17	41	30	127

Figure 6. Pruned bipartite network of art and architecture styles linked via biographies, the latter colored by nationality



not represented in the scholarly version. The rightmost part was occupied by styles representing more recent post-WWII developments in the arts.

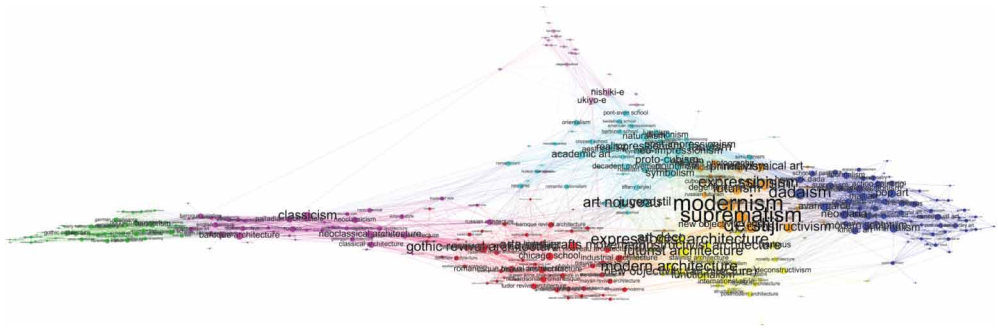
Using the palette from Figure 1 to color the person nodes by nationality again revealed a separate strand of Japanese styles at the bottom which merged with the main strand at around 1900, but also highlighted other styles to be more homogeneous regarding the origin of interlinked persons, such as "Arts and Crafts Movement", "Futurism", "Modernisme", "School of Paris", "Abstract Expressionism" and "Socialist Realism". Especially the latter was separated from the main body and contained many artists not featured in the ULAN, underlining the specific focus of the biographies collected there. Similarly detached and not even directly related to art history, "20th-century classical music" was also marked as art movement in Wikidata and thus included in the visualization. Its connection with the main body reflects the mutual influence between different artistic disciplines and invites for further exploration.

Bipartite Projections

The bipartite person-style network could be reduced to two unipartite networks, one connecting pairs of styles via links weighted by the number of commonly linked persons, the other one vice versa, via bipartite network projection as described in (M. Newman, 2010, p. 123 ff.). Especially the resulting style-style network was of interest to be studied in more detail but as encountered for the nationality network, it featured a few strongly interconnected styles which masked potentially relevant interactions between less prominent styles. This called for another application of a standardized-residual-based filtering technique and the FDSM method described in (Zweig & Kaufmann, 2011) was a suitable approach for bipartite projections. The required 10,000 random bipartite networks could be efficiently calculated using functions from the R library BiRewire, demonstrating that available software components from the domain of Bioinformatics could be straight-forwardly re-used in cultural analytics settings.

The unipartite network of styles weighted by the FDSM-based standardized residuals was subsequently filtered by pruning all style-style links having standardized residual weights below three and its resulting visualization is shown in Figure 7. Clusters of densely interconnected styles

Figure 7. Projection of pruned person-style network between art and architecture styles



were identified via the Louvain clustering method and are shown in different colors in the Figure, they generally appeared to follow a chronological order: From left to right, the clusters started with a group of Romanesque/Gothic/Renaissance arts and architecture styles (green) which blended into a group of Baroque/Classicist styles (purple) whose end at around 1800 again marked the separation of two relatively distinct strand of groups of 19th (red) and 20th/21st century (yellow) architecture styles at the bottom and groups of 19th (cyan), early 20th modernist (orange) and postwar/21st century (blue) art styles at the top. In contrast to the immediate visualization of the person-style network in Figure 6, the representation of the projected and statistically filtered style-style network suggested that modernism represented a period where art and architecture approximated each other after their long phase of separation, just to be followed by more recent developments throughout which they appeared to drift apart again.

RELATED WORK

As far as data-based visualizations of art historical developments are concerned, a main distinction can be drawn between formalist and social-context-oriented conceptions of art historical processes. Approaches towards the former include the clustering of artworks by image features such as dominant color, saturation and brightness, discussed in (Manovich, 2015), while more recent work such as described in (Elgammal, Mazzone, Liu, Kim, & Elhoseiny, 2018) applied contemporary machine-learning methods to extract stylistic relationships between artworks based on Wölfflin's seminal work on stylistic patterns. Potentially located between formalist and social-context approaches, visualizations of metadata collections often contain aspects from both. Visualizing items from the Tate collection metadata, the work presented in (Boyd Davis & Kräutli, 2014) for example discussed the benefits of using visual means to navigate large museum collections across different dimensions. Regarding social context, (Schich et al., 2014) presented visualized patterns of migration via persons featured in the ULAN, providing an interesting view on art historical developments from this perspective. More recent work from the same first author and others presented in (Schich, Huemer, Adamczyk, Manovich, & Liu, 2017) used network visualization to show buyer-seller interactions at different geographic locations derived from the Getty Provenance Index. Visualizations of the ULAN associative link network focusing on ego networks around single persons selected from the ULAN were used in early work in the context of cultural web portals such as CultureSampo, described in (Kurki & Hyvönen, 2009) and for constructing virtual museum landscapes such as described in (Goldfarb, Arends, Froschauer, & Merkl, 2011). Biographical networks on Wikipedia not specifically focused on art history have been discussed and compared across different language versions in works such as presented in (Aragon, Laniado, Kaltenbrunner, & Volkovich, 2012) and in (Eom et al., 2015). While the latter did not use visualizations of the network for their comparison, the former presented

a view on the network of Wikipedia biographies simultaneously present in multiple languages, thus highlighting their agreement and not their differences. Moreover, aspects regarding the chronological structure of the network were not considered. The work presented in (Samoilenko et al., 2016) studied a network of Wikipedia language versions based on commonly featured topics, revealing significant clusters of related language versions appearing to follow socio-cultural commonalities, for which the authors identified bilinguality and shared religion to be the driving factors.

CONCLUSION

This work presented visualizations of biographical networks of important persons from art history as present in the Getty ULAN and in interlinked biographies in different Wikipedia language versions. Their mutual comparison with each other and with existing scholarly examples revealed both similarities based on the common canon of art history as well as differences due to present institutional and cultural biases, the latter especially becoming visible in the varying coverage of nationalities in different Wikipedia language versions. The analysis of the aggregated network of interlinked nationalities revealed high level interactions between cultures which appeared to be clustered by socio-historical boundaries such as colonialism and the changing extent of historical empires. The extraction and visualization of a bipartite network of biographies and art and architecture styles introduced an additional perspective on art history on Wikipedia and enabled to go beyond the ULAN as source for identifying relevant persons. This revealed interesting differences in coverage such as the exclusive presence of many USSR artists in the Russian Wikipedia, for which no corresponding ULAN records could be identified. Overall, the results suggested that large enough numbers of interlinked biographies and descriptions of styles and periods can be used to construct large-scale views on art historical developments, but the available data sources must be handled with care due to identified biases. Initially intended as navigational aids for virtual art collections, such views can thus also be used as tool to reveal the selective coverage of individual cultural heritage datasets. As shown for the analysis of the multi-lingual Wikipedia networks, the culturally specific foci of individual datasets can be leveraged for the good by uniting them in a consolidated dataset representing more than the mere sum of its parts, potentially leading to more global views on cultural developments.

REFERENCES

Agresti, A. (2007). *An Introduction to Categorical Data Analysis*. Hoboken, NJ: Wiley-Interscience. doi:10.1002/0470114754

Aragon, P., Laniado, D., Kaltenbrunner, A., & Volkovich, Y. (2012). Biographical Social Networks on Wikipedia: A Cross-cultural Study of Links That Made History. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration (WIKISYM)*. New York: ACM. doi:10.1145/2462932.2462958

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. (2007). DBpedia: A Nucleus for a Web of Open Data. In K. Aberer et al. (Ed.), *Proceedings of the 6th International Semantic Web Conference* (Vol. 4825). Busan: Springer. doi:10.1007/978-3-540-76298-0_52

Baca, M., & Gill, M. (2015). Encoding Multilingual Knowledge Systems in the Digital Age: The Getty Vocabularies. *Knowledge Organization*, *42*(4), 232–243. doi:10.5771/0943-7444-2015-4-232

Barr, A. (1941). *Advis*ory Committee Report on Museum Collections. New York: The Museum of Modern Art Archive, P9. https://api.moma.org/Images/Size3/Images/MA70_CCCR.jpg

Barr, A. (1986). *Cubism and Abstract Art*. Cambridge, MA: Belknap Press of Harvard University Press. (Original work published 1936)

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics*, *2008*(10), P10008. doi:10.1088/1742-5468/2008/10/P10008

Boyd Davis, S., & Kräutli, F. (2014). *Scholarly Chronographics: Can a Timeline be Useful in Historiography?* Paper presented at the 10th European Social Science History Conference, Vienna, Austria.

Elgammal, A., Mazzone, M., Liu, B., Kim, D., & Elhoseiny, M. (2018). The Shape of Art History in the Eyes of the Machine. In *32. AAAI Conference on Artificial Intelligence*. New Orleans, LA.

Eom, Y.-H., Aragon, P., Laniado, D., Kaltenbrunner, A., Vigna, S., & Shepelyansky, D. L. (2015). Interactions of Cultures and Top People of Wikipedia from Ranking of 24 Language Editions. *PLoS One*, *10*(3), e0114825. doi:10.1371/journal.pone.0114825 PMID:25738291

Goldfarb, D., Arends, M., Froschauer, J., & Merkl, D. (2011). Revisiting 3d Information Landscapes for the Display of Art Historical Web Content. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*. ACM. doi:10.1145/2071423.2071480

Goldfarb, D., Arends, M., Froschauer, J., & Merkl, D. (2013). Comparing Art Historical Networks. *Leonardo*, *46*(3), 279. doi:10.1162/LEON_a_00575

Goldfarb, D., Arends, M., Froschauer, J., Weingartner, M., & Merkl, D. (2014). Collectivizing the Barr Model. *Leonardo*, *47*(3), 270. doi:10.1162/LEON_a_00773

Green-Armytage, P. (2010). A Colour Alphabet and the Limits of Colour Coding. *JAIC - Journal of the International Colour Association, 5*.

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLoS One*, *9*(6), e98679. doi:10.1371/journal.pone.0098679 PMID:24914678

Joyeux-Prunel, B. (2014). The Uses and Abuses of Peripheries in Art History. *Artl@s. Bulletin*, *3*(1).

Klein, M., & Kyrios, A. (2013). VIAFbot and the Integration of Library Data on Wikipedia. *Code4Lib Journal*, 2.

Kurki, J., & Hyvönen, E. (2009). Authority Control of People and Organizations on the Semantic Web. In *Proceedings of the International Conferences on Digital Libraries and the Semantic Web*.

Manovich, L. (2015). Data Science and Digital Art History. *Intl. Journal for Digital Art History,* 1. https://doi.org/10.11588/dah.2015.1.21631

Newman, M. (2010). *Networks: An Introduction*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199206650.001.0001

Newman, M. E. J. (2006, June). Modularity and Community Structure in Networks. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(23), 8577–8582. doi:10.1073/pnas.0601602103 PMID:16723398

Newton, E. (1956). *European Painting and Sculpture* (5th ed.). Harmondsworth, Middlesex: Penguin. (Original work published 1941)

Samoilenko, A., Karimi, F., Edler, D., Kunegis, J., & Strohmaier, M. (2016). Linguistic Neighbourhoods: Explaining Cultural Borders on Wikipedia through Multilingual Co-Editing Activity. *EPJ Data Science*, *5*(1). https://doi.org/10.1140/epjds/s13688-016-0070-8

Schich, M., Huemer, C., Adamczyk, P., Manovich, L., & Liu, Y.-Y. (2017). Network Dimensions in the Getty Provenance Index. *arXiv preprint arXiv:1706.02804*.

Schich, M., Song, C., Ahn, Y.-Y., Mirsky, A., Martino, M., Barabási, A.-L., & Helbing, D. (2014). A Network Framework of Cultural History. *Science*, *345*(6196), 558–562. doi:10.1126/science.1240064 PMID:25082701

Schmidt-Burkhardt, A. (2005). *Stammbäume der Kunst: Zur Genealogie der Avantgarde*. Berlin: Akademie Verlag.

Strass, F. (1828). *Der Strom der Zeiten oder bildliche Darstellung der Weltgeschichte von den ältesten bis auf die neuesten Zeiten*. Leipzig: Köchly. (Original work published 1804), http://mdz-nbn-resolving.de/urn:nbn:de:bvb:12-bsb00003029-3

Vrandečić, D., & Krötzsch, M. (2014). Wikidata: A Free Collaborative Knowledgebase. *Communications of the ACM*, *57*(10), 78–85. doi:10.1145/2629489

Zweig, K. A., & Kaufmann, M. (2011, April). A Systematic Approach to the One-Mode Projection of Bipartite Graphs. *Social Network Analysis and Mining*, *1*(3), 187–218. doi:10.1007/s13278-011-0021-0

Doron Goldfarb received his MS in Computer Science from the Vienna University of Technology in 2008. Since then, he has been involved in different national and European scale research projects in the context of digital cultural heritage and virtual research infrastructures.

Dieter Merkl is Associate Professor of Applied Computer Science at the Institute of Information Systems Engineering, Vienna University of Technology, Austria. His main research interests are in the areas of information retrieval, data mining, and interaction design. He has published more than 140 scientific papers in these areas.