Identifying Vilhelm Hammershøi's Career Phases: A Data-Driven Technical and Art Historical Investigation of Painting Materials and Themes

GIANLUCA PASTORELLI, National Gallery of Denmark/Statens Museum for Kunst (SMK), Copenhagen, Denmark
ANNETTE ROSENVOLD HVIDT, Independent Art Historian, Copenhagen, Denmark
KRISTIAN PRÆSTHOLM, National Gallery of Denmark/Statens Museum for Kunst (SMK), Copenhagen, Denmark
LOA LUDVIGSEN, National Gallery of Denmark/Statens Museum for Kunst (SMK), Copenhagen, Denmark and Kunstfokus, Copenhagen, Denmark
TROELS FILTENBORG and PAULINE LEHMANN BANKE, National Gallery of Denmark/Statens Museum for Kunst (SMK), Copenhagen, Denmark
VERONICA BIOLCATI, National Gallery of Denmark/Statens Museum for Kunst (SMK), Copenhagen, Denmark and University of Paris 1 Panthéon-Sorbonne, Paris, France
ANNE HAACK CHRISTENSEN, National Gallery of Denmark/Statens Museum for Kunst (SMK), Copenhagen, Denmark

This study aimed to combine art historical research, material analysis and computational techniques to identify distinct phases in the Danish artist Vilhelm Hammershøi's career, revealing patterns in his choice of motifs and painting materials, particularly canvases and ground layers, across a large selection of his works. A relational database was developed to organise large amounts of observations and results from the examination of nearly 140 canvas paintings by the artist, including art historical information, structural characteristics of canvases from X-ray imaging, and chemical properties of preparatory layers from scanning electron microscopy coupled with energy dispersive X-ray spectroscopy. A data mining technique based on self-organising maps and hierarchical cluster analysis identified distinct categories of canvas weaves and ground layers, each with a unique structure and pigment composition. At the same time, network analysis based on art historical research, including aspects such as the dates of creation, origins and motifs of the paintings, revealed inter-painting connections that highlight specific groups of works in Hammershøi's oeuvre. The integration of these approaches uncovered significant developments in the artist's material and stylistic choices, defining three career phases with key transitions around 1890

This study was funded by Augustinus Fonden and the New Carlsberg Foundation.
Authors' Contact Information: Gianluca Pastorelli (corresponding author), National Gallery of Denmark/Statens Museum for Kunst (SMK), Copenhagen, Denmark; e-mail: gipa@smk.dk; Annette Rosenvold Hvidt, Independent Art Historian, Copenhagen, Denmark; e-mail: annette.rosenvold@hvidt.net; Kristian Præstholm, National Gallery of Denmark/Statens Museum for Kunst (SMK), Copenhagen, Denmark; e-mail: krpr@smk.dk; Loa Ludvigsen, National Gallery of Denmark/Statens Museum for Kunst (SMK), Copenhagen, Denmark and Kunstfokus, Copenhagen, Denmark; e-mail: loa.ludvigsen@gmail.com; Troels Filtenborg, National Gallery of Denmark/Statens Museum for Kunst (SMK), Copenhagen, Denmark; e-mail: trff@webspeed.dk; Pauline Lehmann Banke, National Gallery of Denmark/Statens Museum for Kunst (SMK), Copenhagen, Denmark; e-mail: PLB@smk.dk; Veronica Biolcati, National Gallery of Denmark/Statens Museum for Kunst (SMK), Copenhagen, Denmark and University of Paris 1 Panthéon-Sorbonne, Paris, France; e-mail: veronica.biolcati1@gmail.com; Anne Haack Christensen, National Gallery of Denmark/Statens Museum for Kunst (SMK), Copenhagen, Denmark; e-mail: AHC@smk.dk.

ACM 1556-4711/2026/6-ART29
https://doi.org/10.1145/3811816

and 1900, likely reflecting influences from his travels in France, Italy and the UK. These findings address a critical gap in art historical research, as no prior studies have systematically examined Hammershøi's artistic evolution through such a comprehensive technical approach. The study also has practical implications for conservation, offering insights into material properties that can aid in assessing the condition of the artist's paintings. In conclusion, this research establishes a valuable foundation for understanding developments in Hammershøi's method, contributing to both art history and preservation practices.

CCS Concepts: • **Applied computing → Digital libraries and archives**; • **Information systems → Data mining**; *Database design and models*; • **Computing methodologies → Machine learning**;

Additional Key Words and Phrases: heritage science, data science, painting collection, machine learning, network analysis, conservation

ACM Reference format:
Gianluca Pastorelli, Annette Rosenvold Hvidt, Kristian Præstholm, Loa Ludvigsen, Troels Filtenborg, Pauline Lehmann Banke, Veronica Biolcati, and Anne Haack Christensen. 2026. Identifying Vilhelm Hammershøi's Career Phases: A Data-Driven Technical and Art Historical Investigation of Painting Materials and Themes. *ACM J. Comput. Cult. Herit.* 19, 2, Article 29 (June 2026), 15 pages.
https://doi.org/10.1145/3811816

1 Introduction

Heritage science is revolutionising the way cultural institutions analyse and interpret large collections of artworks and associated data by leveraging innovative data science tools. These tools integrate systems for data acquisition, storage, processing, exploration, visualisation and communication, ultimately enhancing our understanding of artworks, improving conservation methods and facilitating decision-making in dissemination practices [1]. The intersection of heritage science and data science not only benefits art historical research but also promotes interdisciplinary collaboration, bridging the gaps between art historians, conservators and scientists, and leading to a deeper understanding of our cultural heritage [2]. However, despite these advances, the integration of data science into heritage science still faces significant challenges. Although the fields of conservation science and technical art history have long relied on scientific methods such as technical imaging and materials analysis to study artworks, advancements in these techniques have led to increasingly efficient visual, structural and chemical data acquisition, enabling the rapid collection of vast amounts of heterogeneous data [3]. When applied to large collections of artworks, the analysis of the extensive datasets generated by these methods and the extraction of meaningful information prove highly demanding tasks [4]. As a result, key insights such as patterns, trends, relationships and groupings often remain hidden, limiting the understanding of researchers in art history and heritage science. Therefore, professionals in museums, universities and research institutions require more effective and scalable methods to retrieve, examine and visualise large heritage datasets, in order to uncover the essential knowledge they contain [5]. Furthermore, as cultural institutions increasingly digitise their collections, the ability to efficiently manage and analyse substantial volumes of data becomes progressively crucial.

In response to these emerging challenges, heritage science must continue to evolve, allowing the introduction of new large-scale data science methodologies that can considerably enhance traditional approaches and support art historical research (including technical art history) [6–8]. For example, the systematic study of artists' mobility and materials provenance has become a significant focus in heritage science, and several studies have demonstrated how the integration of traditional scientific methods with innovative data-driven approaches can reveal the connection between artistic production and the availability of materials across different geographical regions [9–11]. In this context, the Vilhelm Hammershøi Digital Archive (ViHDA) project, a 5-year research study launched in 2020 at the National Gallery of Denmark (SMK), offers a compelling example. Carried out by a cross-disciplinary team of paintings conservators, art historians, imaging specialists, heritage scientists and digital experts, the project involved the systematic visual, technical and scientific examination of nearly 140 paintings by the renowned Danish painter Vilhelm Hammershøi (1864–1916) [12, 13]. Among other purposes, the study aimed to identify groups and patterns in Hammershøi's artistic production by correlating chronological, geographical, structural, chemical and stylistic elements. Specifically, the research presented in this article investigated whether Hammershøi's career phases could be identified based on his use of motifs and painting materials, particularly canvases and ground layers, through in-depth analysis of a large volume of historical and scientific data. While some scholars have already attempted to delineate the different phases of Hammershøi's artistic career in terms of motif, style and chronology, their efforts are presented loosely in extended chapters rather than in a structured format [12, 14]. In this study, the term *career phase* refers to distinct periods in which Hammershøi's production exhibits specific patterns in the choice of supports, grounds and themes. Supports and preparatory layers were selected because they are fundamental materials in canvas painting production and can be associated with the artist's technical and thematic evolution over time, thus serving as indicators of career phases. However, a more comprehensive analysis is required to fully characterise the phases of Hammershøi's career and to interpret them as reflections of his intentions; future research will expand on this framework by incorporating additional technical and stylistic variables.

By utilising a multifaceted approach that integrates a relational database of historical, structural and chemical information with unsupervised machine learning methods and network analysis techniques, this study demonstrated how a vast collection of interconnected data could be exploited to complement and enrich art historical and technical interpretations of a large body of the artist's paintings. Additionally, it provided insights into material availability across different periods and geographical locations, highlighting the value of data-driven analysis in the study of artistic practice.

2 Materials and Methods

To determine whether distinct phases in Vilhelm Hammershøi's career could be identified through his use of thematic elements and painting materials, this study combined art historical research, material analysis and computational approaches to reveal patterns in motifs, canvas characteristics and ground layer compositions across a large selection of his works. To address this objective, the study relied on the following methods.

2.1 Data Acquisition

A total of 137 canvas paintings by the artist, held in Danish and international museums as well as private collections with granted access, were investigated, and each was assigned a unique identification number. While the exact number of Hammershøi's paintings on canvas is difficult to determine, as no complete record of his works exists, his total oeuvre is estimated at approximately 400 works, including both drawings and paintings, with the majority executed in oil on canvas [13, 14]. However, this estimate is likely still incomplete, as newly discovered works attributed to him continue to surface, further expanding the known corpus of his oeuvre. The 137 canvas paintings examined in this study are presented in the chronological distribution chart shown in Figure 1, and despite potential biases from limited access to many of the artist's works, the sample is considered fairly representative. Descriptive and art historical data such as dimensions, date and place of production, archival documentation and stylistic assessments of the paintings were manually recorded. For each painting, a number of related works were identified within the investigated set based on similarities in motifs, formats, or time and place of creation when sufficient information was available. At the same time, the materials used in the paintings were characterised by analysing structural and chemical properties related to canvases and ground layers, respectively. Due to logistical and technical constraints, including the lack of radio-opacity in the ground layer and restrictions on sample collection, not all 137 works underwent the full suite of inspections.

To generate quantitative data for assessing canvas characteristics, X-radiography was executed on 89 of the 137 paintings using a Yxlon portable constant potential X-ray unit (SMART Evo 160 D X-ray tube) at 28 to 32 kV,

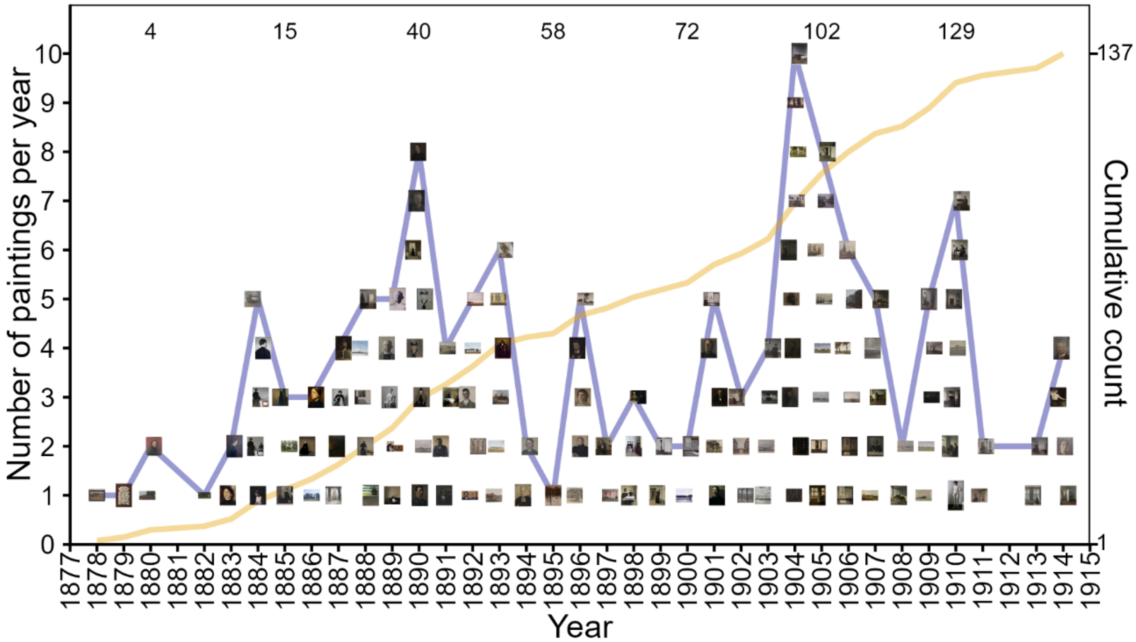

Fig. 1. Chronological distribution of Hammershøi's canvas paintings examined in this study. A connecting blue line illustrates the trend in peak yearly availability, while the orange line represents the cumulative total (cumulative counts for successive 5-year intervals are indicated along the upper margin).

5 mA, with exposure times ranging from 25 to 90 s. The images were captured on $30 \times 40\,\text{cm}^2$ high-definition image plates (HD-IP) and scanned using a Dürr CR tube scanner. The distance between the X-ray source and the film was 110 cm, and a 3-mm AlMg3 filter was employed. Subsequently, the countingVermeer software (threadCount28.cdf, Wolframplayer version 12.3) [15, 16] was used to analyse the X-ray images of the canvases, estimating vertical and horizontal thread densities and angles [17, 18]. A linear filter was applied to minimise noise while preserving the integrity of the weave pattern. Block sizes of either $0.5 \times 0.5\,\text{cm}^2$ or $0.75 \times 0.75\,\text{cm}^2$ were chosen to balance spatial resolution with statistical reliability, ensuring that subtle variations in thread density were captured, and an overlap of 6 units between adjacent blocks was implemented to maintain continuity across the analysed regions. Fast Fourier transforms were computed using sizes of either 512 or 1,024, optimising the tradeoff between computational efficiency and the ability to resolve the periodicity of the weave.

To characterise the pigments and fillers present in the preparatory layers, one or more representative samples of the ground were collected from the edges of 95 of the 137 paintings. All samples were embedded in Technovit 2000 LC light-curing resin from Kulzer Technik (Wehrheim, DE) and prepared as cross sections by polishing the transverse plane. **Scanning Electron Microscopy Coupled with Energy-Dispersive X-Ray Spectroscopy (SEM-EDXS)** was employed to analyse the cross sections. Elemental analysis and mapping were performed using a HITACHI S-3400N scanning electron microscope equipped with a Bruker Quantax 200 energy-dispersive X-ray spectroscopy detection system featuring two Peltier-cooled XFlash silicon drift detectors, each with an active area of $20\,\text{mm}^2$. Measurements were carried out in variable pressure mode (30 Pa) on the non-coated polished sections, using a voltage of 20 kV, a probe current of 50 μA and a working distance of 10 mm. A combination of multi-point analysis and X-ray elemental mapping was employed. Specific areas to be examined for elemental composition with multi-point measurements were carefully selected onto SEM backscattered electron images manually, making sure that the target areas on each layer were representative of the entire layer. Large particles of single pigments were measured individually for pigment characterisation purposes, and were not included in the areas representative of the layers. X-ray elemental mapping was used to visualise the distributions of the elements present in each layer. Acquisition times (live time) were 60 s for the analysis of each selected area and 600 s for the production of the elemental maps (the latter corresponding to a dwell time of approximately 7 ms per measured point). The acquisition and analysis of the EDXS data were performed using Bruker ESPRIT (version 1.9) software. All spectra were subjected to net-intensity (background-corrected peak) fitting, and no additional smoothing or binning was applied in order to retain the full spectral resolution.

2.2 Data Management

A relational database was built to store, link and access the collected historical, structural, and chemical data. The database was designed with a backend, implemented using MySQL Workbench 8.0 CE for data structuring, storage and relational organisation, and a frontend, developed in Microsoft Access (version 2408) for data entry, query execution and retrieval. This architecture enabled the efficient management of a wide-ranging and heterogeneous collection of data, extending beyond the above-mentioned historical documentation and material characterisation to encompass information derived from numerous in-depth painting inspections such as visual examination, multispectral imaging and a broad array of chemical analyses. The resulting database comprised 36 tables, each dedicated to a specific category of data, such as descriptive metadata, visual observations, structural features or chemical properties. A schema incorporating multiple relationships was implemented to ensure data integrity and facilitate complex queries across these interconnected datasets. Therefore, a series of **Structured Query Language (SQL)** statements was employed to select and join relevant data content for this specific study into a single, unified dataset, including all painting records that contained at least two of four designated groups of attributes (Table 1). This approach accounted for variations in examination data, ensuring both fully and partially inspected paintings were included, resulting in a final dataset of 126 records.

2.3 Data Preparation

The dataset generated by the SQL query was imported into the R environment (version 4.5.1) for pre-processing and analysis; all computational tasks were performed on a mid-range laptop PC equipped with a 10th gen Intel Core i7-10610U processor and running Windows 11 Enterprise (version 23H2). The dataset originally consisted of 126 observations, each linked to a unique painting identifier, and 2,472 variables representing descriptive, historical, structural and chemical properties. Notably, the EDXS data accounted for 2,460 variables, corresponding to individual X-ray energy channels between 0.7 and 13.0 keV. For paintings with a production date expressed as a range, the rounded-down midpoint year was calculated and utilised for all chronological analyses. Data preparation involved the creation of three distinct subsets, namely **Canvas Properties (CanPro)**, **Ground Layer Composition (GndCom)** and **Inter-Painting Connections (PntCon)**. Each subset was generated by selecting contextually relevant variables (as detailed in Table 1), and subsequent inspection confirmed the absence of missing values, outliers and duplicate records. For the subset CanPro, two additional variables related to the format chosen by the artist, namely 'canvas size' (width × height) and 'aspect ratio' (width/height), were computed and appended, while the 'weave type' categorical variable was converted into two binary numerical variables using dummy coding. To facilitate comparative analyses, all numerical variables in subsets CanPro and GndCom were min–max normalised to a 0–1 range. Due to incomplete analysis across the 126 paintings, subsets CanPro, GndCom and PntCon resulted in 71, 92 and 88 observations, respectively; despite this reduction, the sample sizes were considered sufficient to support robust conclusions.

2.4 Data Analysis and Visualisation

2.4.1 ***Self-Organising Maps Coupled with Hierarchical Cluster Analysis (SOMHCA)*** *Examination.* To investigate potential associations between paintings based on canvas properties and ground layer compositions,

Table 1. Overview of the Dataset Structure Used for This Study, Showing Attribute Groups, Corresponding Fields, Data Types and Relevant Data Analysis Groups

Group of Attributes	Field/Variable	Data Type	Related Subsets
General information (required)	Unique painting identifier	Integer	CanPro GndCom PntCon
	Year of production (exact year or start year of a range)	Integer	CanPro GndCom PntCon
	Year of production (end year of a range, if applicable)	Integer	CanPro GndCom PntCon
	Country of production	Text string	CanPro GndCom PntCon
	Height (cm)	Numeric	CanPro
	Width (cm)	Numeric	CanPro
	Link to painting image	Text string (hyperlink)	PntCon
Canvas weave properties (if available)	Average horizontal thread density (threads/cm)	Numeric	CanPro
	Average vertical thread density (threads/cm)	Numeric	CanPro
	Average horizontal thread angle (°)	Numeric	CanPro
	Average vertical thread angle (°)	Numeric	CanPro
	Weave type ("plain" or "basket")	Text string	CanPro
Elemental composition of the ground layer (if available)	Sum spectrum net intensities (counts/1,000 s)[a]	Numeric	GndCom
Connections with other paintings (if available)	List of related painting identifiers[b]	Array	PntCon

[a]All EDXS net intensities were recorded in a data matrix composed of 95 observations corresponding to the measured samples, and 2,460 variables associated with the X-ray energies (in the 0.7–13.0 keV range) of the sum spectrum of each sample. When multiple samples were collected and analysed from a single painting, the sample that best represented the overall ground layer composition was selected for inclusion in the final dataset.
[b]Connections between paintings were established by identifying thematic, stylistic and historical similarities through visual evaluation and art historical expertise.

subsets CanPro and GndCom (excluding the variables 'Year of production' and 'Country of production') were independently analysed using an unsupervised machine learning technique, namely SOMHCA. Although the method is described in detail elsewhere [19], we provide a brief overview here. **Self-Organising Maps (SOM)** were used to categorise the high-dimensional data in both subsets while preserving the relative relationships between observations. The optimal SOM grid size was determined through an iterative growing process, evaluating multiple quality measures (i.e., quantisation error, topographic error, Kaski-Lagus error, percentage of explained variance, and various combinations thereof), and the SOM model was then trained using the grid size that maximised performance. **Hierarchical Cluster Analysis (HCA)** was applied to the SOM nodes using the **Kelley-Gardner-Sutcliffe (KGS)** penalty function to determine the optimal number of clusters. Finally, the assigned clusters were mapped back to the original subsets for visualisation and interpretation of patterns that may correspond to different periods of the artist's career. This integrated approach offers more interpretable and accurate results compared to standard techniques like k-means or **Principal Component Analysis (PCA)**-based clustering [19]. The SOMHCA method was implemented using the *somhca* R package (version 0.3.0) [20], which includes functions for data preparation (*loadMatrix()*), grid size optimisation (*optimalSOM()*), training of the SOM model (*finalSOM()*), visualisation of feature distributions (*generatePlot()*), clustering of the SOM nodes (*clusterSOM()*) and cluster assignment retrieval (*getClusterData()*). Clustering stability was assessed by repeating the SOMHCA pipeline 20 times with different random seeds, while keeping the optimal grid size and optimal number of clusters fixed.

2.4.2 Network Analysis. To refine and contextualise the findings from the analysis of subsets CanPro and GndCom, and integrate them with art historical information, the data of subset PntCon was transformed into a network where each painting was represented as a node and the inter-painting connections defined the edges [21]; all connections were treated equally without additional weighting. Node positions were determined using a force-directed layout algorithm implemented via the *igraph* package (version 2.1.4) [22]. Painting images corresponding to the nodes were incorporated into the network visualisation to explore groups of thematic elements. Additional numeric and categorical metadata, including chronological (year of production) and geographical (place of production) information as well as cluster assignments from the above-described SOMHCA analyses (when available), were merged with the network data by appending them to each painting's record for identifying groups of nodes based on those specific attributes. Since edge direction was not essential for grouping, the network was converted to undirected, ensuring that groups reflected connectivity patterns without influence from individual edge directions. The quantification of dissimilarities between nodes was achieved using Gower's distance [23], enabling the integration of mixed data types and the handling of missing values. An HCA approach was then applied using Ward's method, which minimises the total within-cluster variance at each step to form compact, homogeneous groups [24]. The optimal number of clusters was determined by evaluating the clustering structure using the KGS penalty-based approach, selecting the partition that minimised the clustering cost [25]. The resulting clusters were assigned to the nodes, and an additional graphical representation of the network with nodes coloured by cluster was generated for further visualisation and interpretation of the results.

3 Results and Discussion

This research explored the possibility of distinguishing phases in Hammershøi's career by examining his choice of motifs and painting materials through a comprehensive analysis of extensive historical and scientific data. Chi-square tests were performed to evaluate whether data reduction in the CanPro and GndCom subsets was associated with year or country of production, with significance assessed at $\alpha = 0.05$. None of the tests was statistically significant (CanPro vs. year: $p = 0.11$; CanPro vs. country: $p = 0.079$; GndCom vs. year: $p = 0.20$; GndCom vs. country: $p = 0.37$), indicating that missingness is unlikely to be systematic. These results suggest that the primary conclusions presented in the following sections are not strongly influenced by uneven data coverage. Further details about the specific paintings mentioned in the following sections (Figure S1), along with stability and sensitivity analyses of clustering results assessed via validity indices across different random initializations and numbers of clusters [26, 27] (Figure S2), are available in the supplementary material.

3.1 SOMHCA Examination

An advanced analysis was conducted to identify clusters of canvas characteristics and ground compositions, aiming to determine potential relationships between structural and chemical properties across the studied paintings. Regarding the canvas analysis, hierarchical clustering applied to a 6×6 SOM map identified four clusters with clearly defined characteristics across key variables (Figure 2). Cluster 1a, and to a lesser extent cluster 2a, are the most abundant and display nearly identical weave properties in terms of average thread density

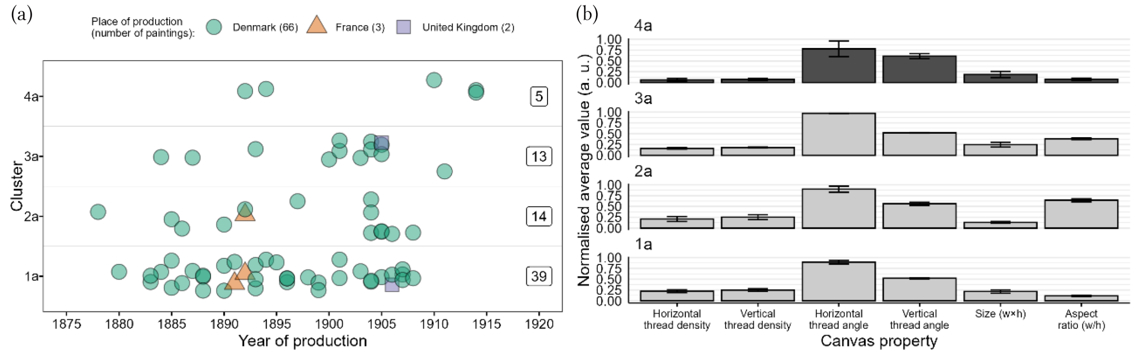

Fig. 2. Distribution of the investigated paintings based on age, geographical origin and canvas properties. (a) Diagram showing paintings categorised by year and place of production, with cluster attribution based on analysis of canvas properties; the number of paintings in each cluster is indicated on the right side of the plot. (b) Bar chart displaying the normalised average values of canvas structural and format properties for each cluster; error bars represent standard errors, and the darker shading of the bars for cluster 4a indicates the basket weave type, in contrast to the plain weave typical of the other clusters.

and thread angle. However, they differ in the format chosen by the artist. Cluster 1a consists predominantly of portrait-format paintings, including two works created in France (a self-portrait and a depiction of a Greek relief from the Louvre) and one painted in the United Kingdom (a London cityscape). In contrast, cluster 2a primarily consists of landscape- and square-format paintings, including a double portrait of the artist and his wife, Ida Hammershøi (née Ilsted), created in Paris during their honeymoon. Notably, these two clusters also contain paintings with some of the highest canvas densities, such as two studies for a portrait of Ida Hammershøi and a view of the Grønsund strait (Falster, Denmark), all dated to 1890. Cluster 3a is characterised by a slightly lower average weave density than the previous clusters and includes another cityscape painted in London. Cluster 4a only contains five portrait-format paintings produced after 1890, three of which were painted toward the end of the artist's career; the canvases in this cluster are distinguished by the lowest average weave density and a basket weave structure.

Concerning the preparatory layers analysis, hierarchical clustering applied to a 6×6 SOM map initially identified five clusters; however, closer inspection revealed that two were characterised by very similar mean spectra. After merging these into a single group by recomputing the analysis with four clusters, the validity indices showed only a minor decrease (Figure S2 in the supplementary material), while the overall interpretability improved, with the resulting mean spectra indicating clearly distinct chemical compositions (Figure 3). To evaluate the potential impact of high dimensionality and channel collinearity, PCA was applied to the centred and scaled spectral matrix. Components explaining 99% of the total variance (77 PCs) were retained, and the SOMHCA workflow was repeated on the reduced dataset. Most validation indices decreased compared to those obtained from the full spectral data (Figure S2 in the supplementary material), indicating that dimensionality reduction affects cluster structure and that chemically relevant information is not confined to the highest-variance components. The majority of the paintings (cluster 1b) have ground layers primarily composed of lead white, which may have been combined with varying amounts of other pigments such as zinc white and calcium carbonate. These additional pigments may have been mixed with the lead white or used separately in double grounds. In some cases, the lead white grounds are mixed with aluminosilicates, likely clay minerals, particularly in works created before 1890. A smaller group of paintings features ground layers predominantly (though not exclusively) composed of either calcium carbonate or zinc white, corresponding to cluster 2b and cluster 3b, respectively. While the zinc-rich ground is present only in paintings executed before 1890, the calcium-rich ground is found exclusively

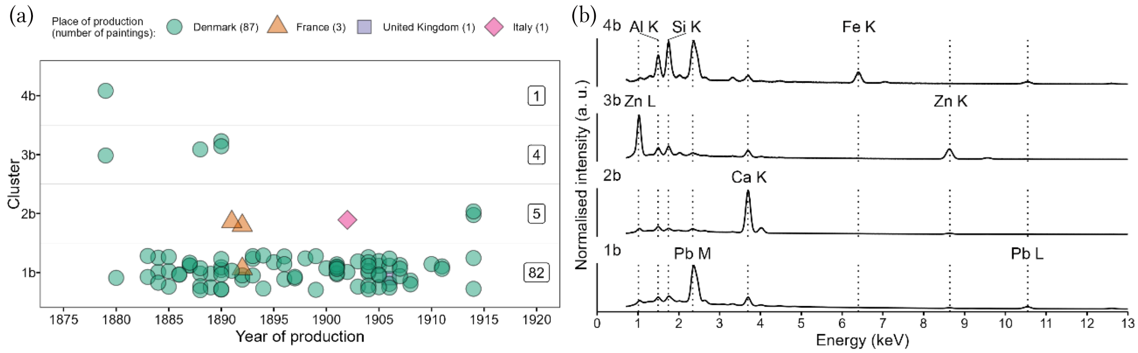

Fig. 3. Distribution of the investigated paintings based on age, geographical origin and ground composition. (a) Diagram showing paintings categorised by year and place of production, with cluster attribution based on the chemical composition of the ground layers; the number of paintings in each cluster is indicated on the right side of the plot. (b) Average EDXS spectra representing the elemental composition of the ground type for each cluster.

in paintings dated after 1890, including the above-mentioned self-portrait and double portrait created in France as well as the only painting in the dataset created in Italy, a 1902 interior of the church of Santo Stefano Rotondo in Rome. Cross-section analysis showed that the ground layer of this church interior consists of a double layer with a unique stratigraphy: a thick lower layer composed mostly of calcium carbonate with a small amount of zinc white, and a thin upper layer of lead white. Further analysis using X-ray fluorescence spectroscopy and fibre optic reflectance spectroscopy revealed that a small number of paintings in cluster 2b contain pigments in the paint layer that appear less common in Hammershøi's typical palette. Notably, ultramarine blue and madder lake were identified in the double portrait painted in Paris (although these observations must be considered preliminary, as systematic analyses for ultramarine and organic colourants have not been performed on all his works), while cobalt yellow was detected in the church interior painted in Italy. Finally, a single painting from the earliest phase of the artist's career has a unique double ground composition, consisting of a lower earth pigment layer (rich in iron and aluminosilicates) combined with an upper lead white layer. This painting is the only one classified in cluster 4b, setting it apart from the other works in the dataset (while several of these also feature double grounds, they are composed exclusively of white pigments). This difference in technique may be explained by the fact that this work dates from the artist's time at the Royal Danish Academy of Fine Arts, where materials were likely provided for academic exercises.

The analysis of the combined canvas and ground clusters, limited to paintings for which both datasets are available, using Fisher's exact test ($p = 0.69$), revealed no significant association between the explored structural and chemical properties (Table 2). This suggests that, given Hammershøi's use of commercially prepared canvases (rather than self-prepared ones, which would have been atypical at the time), the manufacturing of artists' canvases was divided into distinct processes, with the fabric produced industrially in a weaving mill and the ground applied separately in a priming workshop [28]. The majority of paintings (61 works) feature plain-weave canvases with lead white-based grounds. However, four other paintings with a lead white ground were executed on basket weave canvases; these works were produced in two distinct periods, with the first two including portraits of the artist's brother (1892) and mother (1894), and the remaining two dated to 1910 and 1914. Among the paintings with a calcium carbonate-based ground, the two portraits created in Paris share similar plain-weave canvases but differ in format: the self-portrait adheres to what is generally considered a typical portrait format, while the double portrait has the largest aspect ratio among the investigated paintings, with a width nearly twice its height. In contrast, only one painting, a 1914 portrait of Karen Malm, was produced on a basket weave; interestingly, the ground layer of this painting contains calcium sulfate in addition to calcium

Table 2. Contingency Table Showing the Painting Counts by Canvas Structure and Ground Layer Composition

Ground Composition	Canvas Structure				Total
	Cluster 1a	Cluster 2a	Cluster 3a	Cluster 4a	
Cluster 1b	37	12	12	4	65
Cluster 2b	1	1	-	1	3
Cluster 3b	-	-	-	-	-
Cluster 4b	-	-	-	-	-
Total	38	13	12	5	68

Only two of the identified ground clusters include paintings with counterparts across the four canvas clusters, resulting in a limited association between ground layer composition and canvas structure.

carbonate. In general, the year 1890 emerged as a key turning point in Hammershøi's career, marking the adoption of new support and priming materials, such as basket weave canvases and purer lead white grounds, and a transition from zinc white to calcium carbonate grounds as alternatives to lead white. This suggests a shift in material choices or a change in availability after that year, possibly influenced by his well-documented travels [13]. Although only a few of his paintings created outside Denmark are known, it can be inferred that Hammershøi began sourcing materials on-site rather than bringing them with him, possibly reflecting an adaptation to local resources or practical constraints. Alternatively, this change in working habits could be linked to either new supplies and/or suppliers, or a natural inclination to experiment as his artistic approach evolved.

3.2 Network Analysis

A network analysis of Hammershøi's paintings based on art historical information produced an undirected graph that maps the interconnections among his works (Figure 4(a)). This visualisation shows groups of closely related paintings, highlighting potential series or thematic progressions, and identifies key paintings with multiple connections that may serve as stylistic or compositional reference points. For example, one group of paintings focuses on sunlit room interiors (group I), another revolves around Danish landscapes (group II), and a further group is dedicated to depictions of back-turned figures set against various interiors (group III). Other groups centre on portraits of Ida Hammershøi, two of which served as models for a final 'engagement painting' (group IV), as well as his enigmatic, mythological painting *Artemis* in its final form along with two paintings that served as an inspiration and a study piece, respectively (group V). Moreover, the spatial distribution within the graph suggests that these distinct groupings may align with specific historical periods of Hammershøi's oeuvre, potentially reinforcing the previous technical analysis assessments. This assumption could be confirmed by combining this data with the results from the SOMHCA analysis described in Section 3.1, highlighting three clusters of works based on shared time of creation, provenance, motifs and material attributes; although five clusters were initially identified, one containing only three paintings produced in France and another with two paintings produced in the UK were each merged with their respective nearest neighbouring clusters through a subsequent recomputation, resulting in three clusters with a moderate reduction in validity indices but greater interpretability. The generated diagram (Figure 4(b), and the associated Figure S3 in the supplementary material) displays groups that are homogeneous in their characteristics, such as groups III and IV, which feature portraits and figures painted in Denmark before (and including) 1890, all painted on plain-weave canvases mostly primed with lead white grounds. On the other hand, some groups are more varied. For instance, group I includes an interior painted

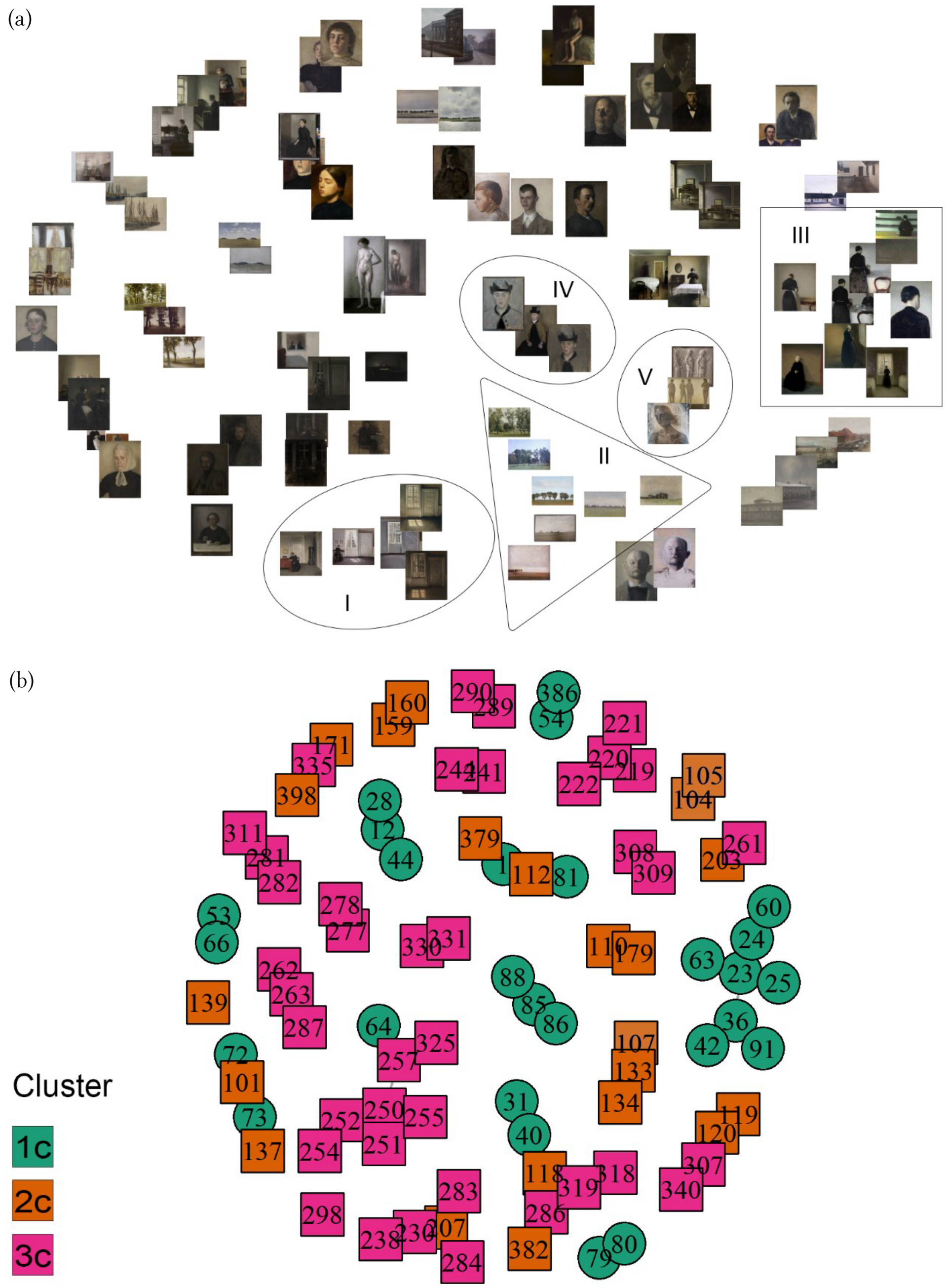

Fig. 4. (a) Network visualisation of 88 paintings by Hammershøi, showing groupings based on art historical information and stylistic assessments; five example groups are outlined and labelled I–V. (b) Similar network visualisation with nodes numbered according to their corresponding unique painting identifiers and coloured to represent combined clusters, which integrate chronological and geographical information along with cluster assignments from the SOMHCA analyses. Closely positioned nodes are connected, while separations between node groups indicate the absence of connections.

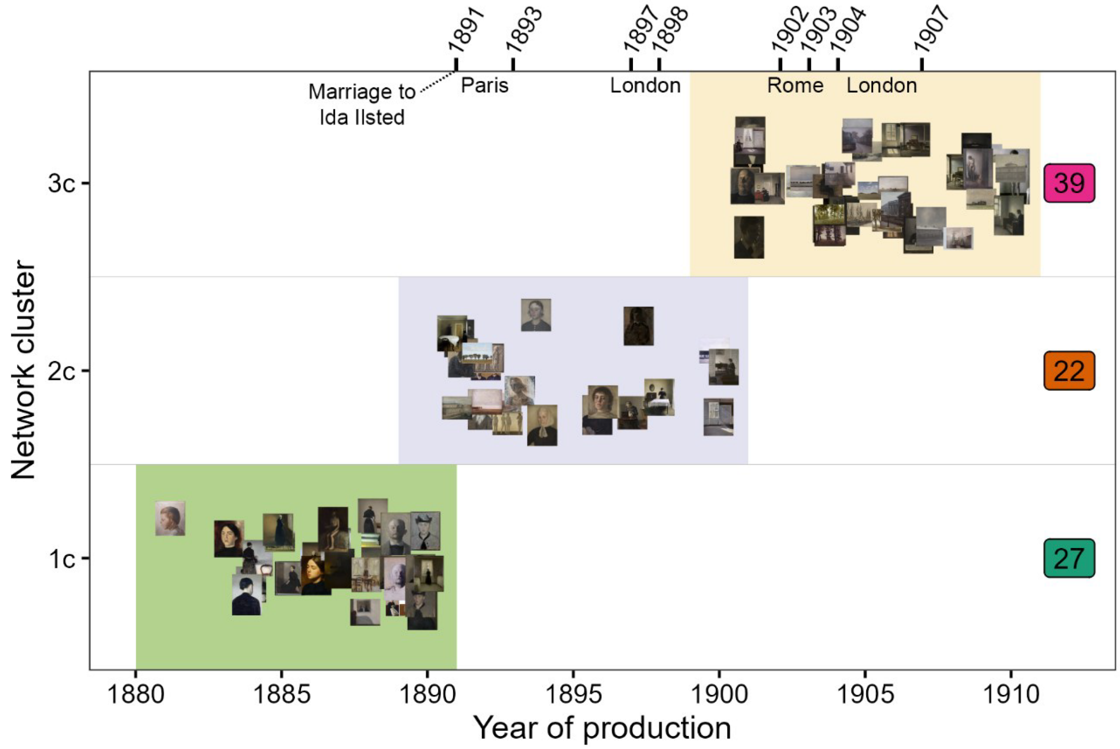

Fig. 5. Timeline of Hammershøi's paintings categorised by production phase and network cluster attribution; the number of paintings in each cluster is indicated on the right side of the plot, with colours matching those used in the network visualisation. Selected events in the artist's life that are relevant to this study, including his marriage and periods spent outside Denmark, are indicated along the upper margin.

on a more lightly woven canvas compared to the others in the same series, while group II consists of landscapes created in Denmark at various times before, between and after his travels in France, Italy and the UK.

When mapped against a timeline, this integrated dataset produced a visualisation that illustrates Hammershøi's artistic phases and potential shifts in choice of materials, stylistic techniques or subject matter influenced in part by geographic context (Figure 5). Based on the paintings included in this project, as the entire oeuvre has not been studied, this visualisation reveals three distinct phases in his artistic production, each spanning approximately a decade. The first phase (*circa* 1880–1890) is characterised by a predominance of portraits of his family, fellow artists and other figures, with only a few landscapes and interiors; these works were executed on dense plain-weave canvases with lead white-based, and occasionally zinc white-based, ground layers. The second phase (*circa* 1890–1900), which includes select works created in France, remains dominated by portraiture featuring family, friends and other figures, as well as his monumental work *Artemis*; this phase also includes some landscapes, cityscapes and interiors, and marks the introduction of new materials such as less dense basket weave canvases, calcium carbonate grounds and purer lead white grounds. In the third phase (*circa* 1900–1910), portraits, interiors and landscapes (along with cityscapes and seascapes) become more evenly distributed, accompanied by the continued use of plain-weave canvases and high-quality lead white grounds. Based on the results from the SOMHCA analysis of paintings not included in the network analysis, two additional phases may be identified: one preceding 1880 and another following 1910, corresponding to the beginning and end of Hammershøi's career, during which, based on the paintings included in this research, the artist primarily painted works in portrait format. The earlier phase is marked by a single case of iron-rich ground use, linked to his studies at the Royal Danish Academy of Fine Arts and absent in the rest of his production, while the later phase features the use of basket weave canvases, one of which is prepared with a calcium carbonate ground, likely remnants from his earlier travels. These findings support the hypothesis that Hammershøi's artistic methodology involved iterative refinements and thematic revisitations, shaped, although not necessarily determined, by factors such as material constraints, availability and improvements.

4 Conclusion

This study demonstrated that distinct phases of Vilhelm Hammershøi's career could be identified by correlating historical records with data related to the artist's stylistic choices and selection of painting materials. A comprehensive relational database was designed to integrate chronological, geographical and technical information, including the structure of canvases and the chemical composition of ground layers, to support an in-depth analysis of his works. By employing data-driven methods such as unsupervised machine learning and network analysis, this study examined a vast collection of interconnected data to enhance art historical interpretations of an extensive selection of the artist's paintings. Based on the paintings included in this project, the research revealed distinct clusters in canvas characteristics, ground layer compositions and thematic elements that correspond to different periods of Hammershøi's artistic production. Notably, transitions around 1890 and 1900 mark key turning points, during which the artist adopted new canvas and ground materials, and shifted from predominantly portraiture to a broader range of subjects. These changes appear to reflect both an evolution in his material and stylistic preferences and adaptations to new working environments during his travels. By merging historical context with scientific data, these findings fill a considerable gap in the existing literature, as no previous studies have combined these approaches to explore Hammershøi's artistic production and technical progress. This innovative methodology, linking heritage science with data science, not only deepens our understanding of his work but also provides a framework for future art historical investigations. The implications of this research also extend to the field of conservation, as by mapping the specific materials and techniques used throughout Hammershøi's career, conservators may gain a valuable tool for assessing the condition of his paintings. A precise understanding of the original materials allows for more informed decisions regarding preservation and restoration, potentially serving as a model for similar studies on other contemporary artists. However, the study has some limitations. The sample size was constrained because not all of the 137 canvas paintings originally included in the database underwent the full suite of analyses, which may affect the overall representativeness of the results. Moreover, the investigation focused exclusively on canvases and ground layers, omitting the systematic analysis of underdrawings and paint layers. While the findings are robust, they might not fully capture the scope of Hammershøi's material practices. Current research is building on these findings by incorporating multispectral imaging, extending chemical analyses to include the paint layers, and examining additional art historical context, such as potential influences from suppliers. These efforts, which require additional interpretation before definitive conclusions can be drawn, are expected to provide a more comprehensive view of Hammershøi's techniques and material evolution, and will be addressed in a separate publication. In addition, all data collected during this research is currently publicly available through an open-access digital archive (hammershoi.smk.dk). This archive features thousands of high-resolution downloadable images and an image viewer designed to allow the comparison of technical images and chemical maps, thus supporting scholarly research and practical conservation efforts, significantly enhancing our understanding of this artist's work.

Acknowledgements

The authors are grateful to Oscar Holm (National Gallery of Denmark) for technical assistance in the preparation of the samples used in this research, and the Royal Danish Academy—Institute of Conservation for the use of the

SEM-EDXS instrumentation. The authors also thank independent researcher Jørgen Wadum (WATS—Wadum Art Technological Studies, Denmark) for his valuable insights.

References

[1] A. Albuerne, J. Grau-Bove, and M. Strlic. 2018. The role of heritage data science in digital heritage. In *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection*. M. Ioannides, E. Fink, R. Brumana, P. Patias, A. Doulamis, J. Martins, and M. Wallace (Eds.), Lecture Notes in Computer Science, Vol. 11196, Springer International Publishing, 616–622. DOI: https://doi.org/10.1007/978-3-030-01762-0_54

[2] L. Manovich. 2015. Data science and digital art history. *International Journal for Digital Art History* 1 (2015), 10–35. DOI: https://doi.org/10.11588/DAH.2015.1.21631

[3] J. Riu and B. Giussani. 2024. Analytical chemistry meets art: The transformative role of chemometrics in cultural heritage preservation. *Chemometrics and Intelligent Laboratory Systems* 247 (Apr. 2024), 105095. DOI: https://doi.org/10.1016/j.chemolab.2024.105095

[4] M. Cardinali. 2019. Digital tools and technical views: The intersection of digital art history and technical art history in a digital archive on the painting technique of Caravaggio and his followers. *Visual Resources* 35, 1–2 (Apr. 2019), 52–73. DOI: https://doi.org/10.1080/01973762.2019.1555351

[5] F. Pozzi and C. H. Stephens. Advances in analytical methods for cultural heritage. *Applied Sciences* 14, 17 (Aug. 2024), 7587. DOI: https://doi.org/10.3390/app14177587.

[6] A. Wasielewski. 2023. *Computational Formalism: Art History and Machine Learning*. The MIT Press. DOI: https://doi.org/10.7551/mitpress/14268.001.0001

[7] G. Castellano and G. Vessio. A deep learning approach to clustering visual arts. *International Journal of Computer Vision* 130, 11 (Nov. 2022), 2590–2605. DOI: https://doi.org/10.1007/s11263-022-01664-y

[8] H. Y. D. Sigaki, M. Perc, and H. V. Ribeiro. 2018. History of art paintings through the lens of entropy and complexity. *Proceedings of the National Academy of Sciences of the United States of America* 115, 37 (Sept. 2018), E8585–E8594. DOI: https://doi.org/10.1073/pnas.1800083115

[9] M. Albrecht, O. de Noord, S. Meloni, A. van Loon, and R. Haswell. Jan Steen's ground layers analysed with principal component analysis. *Heritage Science* 7, Article 53 (Dec. 2019). DOI: https://doi.org/10.1186/s40494-019-0295-5

[10] A. Luther. 2020. Digital provenance, open access, and data-driven art history(1st ed.). In *The Routledge Companion to Digital Humanities and Art History*. K. Brown (Ed.), Routledge, 448–458. DOI: https://doi.org/10.4324/9780429505188-38

[11] J. S. Lozano. 2023. Crunching the canon. *Art History* 46, 1 (Feb. 2023), 197–200. DOI: https://doi.org/10.1111/1467-8365.12701

[12] P. Vad and K. Tindall. 1992. *Vilhelm Hammershøi and Danish Art at the Turn of the Century*. Yale University Press, London.

[13] A. R. Hvidt and G. Oelsner. 2018. *Vilhelm Hammershøi: på Sporet af Det Åbne Billede*. Lindhardt og Ringhof, København.

[14] S. Michaëlis and A. Bramsen. 1918. *Vilhelm Hammershøi: Kunstneren og Hans Værk*. Gyldendalske Boghandel, Nordisk Forlag, København & Kristiania.

[15] W. A. Sethares, Jr. 2017. Computer-assisted manual thread marking. In *Counting Vermeer*. C. R. Johnson (Ed.), RKD Monograph. Retrieved January 28, 2025 from https://countingvermeer.rkdstudies.nl/4-computer-assisted-manual-thread-marking/

[16] W. A. Sethares, Jr. 2017. Automated creation of weave and angle maps. In *Counting Vermeer*. C. R. Johnson (Ed.), RKD Monograph. Retrieved March 12, 2025 from https://countingvermeer.rkdstudies.nl/5-automated-creation-of-weave-maps/

[17] R. C. Johnson, D. H. Johnson, N. Hamashima, H. Sung Yang, and E. Hendriks. On the utility of spectral-maximum-based automated thread counting from X-radiographs of paintings on canvas. *Studies in Conservation* 56, 2 (Mar. 2011), 104–114. DOI: https://doi.org/10.1179/sic.2011.56.2.104

[18] D. H. Johnson, C. Richard Johnson, and R. G. Erdmann. Weave analysis of paintings on canvas from radiographs. *Signal Processing* 93, 3 (Mar. 2013), 527–540. DOI: https://doi.org/10.1016/j.sigpro.2012.05.029

[19] G. Pastorelli, A. S. Ortiz Miranda, and A. H. Christensen. Interpretation of x-ray spectral data using self-organising maps and hierarchical clustering: A study of Vilhelm Hammershøi's use of painting materials. *X-Ray Spectrometry* 53, 5 (Sept. 2024), 392–404. DOI: https://doi.org/10.1002/xrs.3388

[20] G. Pastorelli. 2025. SOMHCA: Self-Organising Maps Coupled with Hierarchical Cluster Analysis. Retrieved from https://CRAN.R-project.org/package=somhca

[21] D. König. 1990. Theory of finite and infinite graphs. In *Theory of Finite and Infinite Graphs*. Birkhäuser Boston, Boston, MA, 45–421. DOI: https://doi.org/10.1007/978-1-4684-8971-2_2.

[22] G. Csárdi, T. Nepusz, K. Müller, S. Horvát, V. Traag, F. Zanini, D. Noom. 2025. *Igraph for R: R Interface of the Igraph Library for Graph Theory and Network Analysis*. Zenodo. DOI: https://doi.org/10.5281/ZENODO.7682609

[23] J. C. Gower. 1971. A general coefficient of similarity and some of its properties. *Biometrics* 27, 4 (Dec. 1971), 857–871. DOI: https://doi.org/10.2307/2528823

[24] J. H. Ward. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 58, 301 (Mar. 1963), 236–244. DOI: https://doi.org/10.1080/01621459.1963.10500845

[25] L. A. Kelley, S. P. Gardner, and M. J. Sutcliffe. 1996. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. *Protein Engineering, Design & Selection* 9, 11 (1996), 1063–1065. DOI: https://doi.org/10.1093/protein/9.11.1063

[26] M. Halkidi, Y. Batistakis, and M. Vazirgiannis. 2001. On clustering validation techniques. *Journal of Intelligent Information Systems* 17, 2–3 (Dec. 2001), 107–145. DOI: https://doi.org/10.1023/A:1012801612483

[27] A. M. Ikotun, F. Habyarimana, and A. E. Ezugwu. 2025. Cluster validity indices for automatic clustering: A comprehensive review. *Heliyon* 11, 2 (Jan. 2025), e41953. DOI: https://doi.org/10.1016/j.heliyon.2025.e41953

[28] C. K. Andersen, T. Filtenborg, M. Scharff, and A. B. Scharff. 2009. The industrialization of canvas production in Denmark and its implications for the preservation of Danish nineteenth century paintings. In *Proceedings of the Incredible Industry, Preserving the evidence of industrial society, Nordisk Konservatorforbund*, 39–49. DOI: https://doi.org/10.13140/2.1.1143.7768

Received 11 April 2025; revised 26 February 2026; accepted 18 April 2026