

Nutzung visueller Analysen für die Erforschung von Netzwerken im Bereich der kunsthistorischen Forschung

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Wirtschaftsinformatik

eingereicht von

Adam Nedas, BSc

Matrikelnummer 01525129

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Univ.Prof. Mag.rer.soc.oec. Dr.rer.soc.oec. Silvia Miksch

Mitwirkung: Dipl.-Ing. Velitchko Filipov, BSc

Wien, 27. November 2022

Adam Nedas

Silvia Miksch



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.



Utilizing Visual Analytics for Network Exploration in the Domain of Art History Research

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Business Informatics

by

Adam Nedas, BSc

Registration Number 01525129

to the Faculty of Informatics

at the TU Wien

Advisor: Univ.Prof. Mag.rer.soc.oec. Dr.rer.soc.oec. Silvia Miksch

Assistance: Dipl.-Ing. Velitchko Filipov, BSc

Vienna, 27th November, 2022

Adam Nedas

Silvia Miksch



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Erklärung zur Verfassung der Arbeit

Adam Nedas, BSc

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 27. November 2022

Adam Nedas



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Danksagung

Ich möchte mich bei Univ.-Prof. Mag. Dr. Silvia Miksch sowie Projektass.(FWF) Dipl.-Ing. Filipov Velitchko, BSc. für ihre äußerst zuverlässige sowie rundum exzellente Betreuung bedanken.

Meine Dankbarkeit gilt auch meinen FreundInnen sowie KollegInnen für ihre Hilfe während meiner Studienlaufbahn.

Abschließend möchte ich meiner Familie danken. Ich hätte diesen Grad nicht ohne ihre Unterstützung erlangen können.



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Acknowledgements

I would like to thank my supervisors Univ.-Prof. Mag. Dr. Silvia Miksch as well as Projektass.(FWF) Dipl.-Ing. Filipov Velitchko, BSc. for their highly reliable and overall excellent support.

My gratitude also goes to my friends and colleagues for their help and guidance throughout my studies.

Finally, I would like to thank my family. I could not have completed this degree without their support.



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Kurzfassung

In der kunsthistorischen Forschung kann die Untersuchung von sozialen Netzwerken Einblicke in komplexe Interaktionen zwischen Künstlerinnen und Künstlern gewähren. Trotz erfolgreichem Einsatz von Netzwerkvisualisierungen im Bereich der Kunstgeschichte gibt es allerdings noch viele offene Fragen sowie Herausforderungen, welche zu lösen sind, um Kunsthistorikerinnen und Kunsthistoriker bestmöglich bei ihrer Forschung zu unterstützen. So gibt es bei den existierenden Ansätzen nur begrenzt Möglichkeiten, kunsthistorische Netzwerke mit Berücksichtigung auf ihren geographischen Bezug sowie ihrer zeitlichen Entwicklung visuell zu erforschen. Des Weiteren wird üblicherweise auf sogenannte Knoten-Link-Diagramme bei der visuellen Darstellung von Netzwerken zurückgegriffen, welche in Vergangenheit im Vergleich mit anderen Darstellungen bei verschiedenen Untersuchungen Schwächen offenbarten. Zuletzt gestaltet es sich schwierig, Anwendungen in Bezug auf ihre Tauglichkeit in der kunstgeschichtlichen Domäne zu evaluieren, da etablierte Evaluierungsverfahren oftmals nicht mit Domänenexpertinnen und Domänenexperten durchgeführt werden.

Wir stellen *Exhibitions Explorer* vor, eine Lösung, welche die interaktive, visuelle Erforschung kunsthistorischer Netzwerke ermöglicht. Die Lösung integriert unterschiedliche Visualisierungsansätze aus der Forschung in ein neues Visualisierungskonzept, welches die Anforderungen der Domäne in den Mittelpunkt stellt. Mit Hilfe eines einsicht-basierten Evaluierungsansatzes zeigen wir, dass, ein passendes Visualisierungskonzept vorausgesetzt, Netzwerkvisualisierungen zu einer hohen Quantität und Qualität an domänenrelevanten Einsichten führen können.



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Abstract

In art historical research, the study of social networks can provide insights into complex interactions between artists. However, despite the successful use of network visualizations in the field of art history, there are still many open questions as well as challenges that need to be solved in order to support art historians in the best possible way for their research. For example, existing approaches offer only limited possibilities to visually explore art historical networks with regard to their geographical context as well as their temporal development. Furthermore, so-called node-link diagrams are usually used for the visual representation of networks, which in the past revealed weaknesses in comparison with other representations in various studies. Finally, it is difficult to evaluate applications with respect to their suitability in the art history domain, since established evaluation approaches are often not conducted with domain experts.

We present *Exhibitions Explorer*, a solution which enables the interactive, visual exploration of art historical networks. The solution integrates different visualization approaches from research into a new visualization concept that focuses on the requirements of the domain. Using an insight-based evaluation approach, we also demonstrate that, provided a suitable visualization concept, network visualizations can lead to a high quantity and quality of domain-relevant insights.



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Contents

Kurzfassung	xi
Abstract	xiii
Contents	xv
1 Introduction	1
1.1 Research Question	2
1.2 Expected Results and Contributions	3
1.3 Structure and Methodology	3
2 Basic Concepts in Graph Theory and Network Science	5
2.1 Foundations in Graph Theory	5
2.2 Network Science	6
3 Related Work	11
3.1 Research Methodology for the Literature Review	11
3.2 Network Analysis and Visualization in Digital Art History Research	13
3.3 State-of-the-Art Network Visualization Techniques	20
3.4 Discussion	28
4 Prototype Design	31
4.1 Data	32
4.2 Users	35
4.3 Tasks	36
4.4 Summary and Resulting Visualization Layouts	38
5 Implementation	39
5.1 Labeled Property Graph Model	39
5.2 Application Architecture	40
5.3 Implementation Details and Approach	41
6 Visualization Design and Use Cases	45
6.1 Map	46
	xv

6.2	Network Creation	48
6.3	Dashboard	50
6.4	Matrix-based Visualization	54
6.5	Node-Link Diagram	57
6.6	Use Cases	59
6.7	Revisiting the Challenges identified in the SLR	60
7	Evaluation	63
7.1	Introduction to the Insight-Based Methodology	63
7.2	Evaluation Protocol	66
7.3	Results	72
8	Discussion	83
8.1	Revisiting the Sub-Questions of this Thesis	83
8.2	Usability Issues in Exhibitions Explorer	86
8.3	Challenges Regarding the Evaluation Protocol and Recommendations for Future Adaptions	87
9	Conclusion	89
9.1	Future Work	89
9.2	Final Remarks	90
	List of Figures	93
	List of Tables	95
	Bibliography	97
	Appendix	109
	Insights	109
	Comments during Evaluation	109
	Post-evaluation Interviews	112

Introduction

The *Database of Modern Exhibitions* [DoMb] contains information about modern art exhibitions in the beginning of the 20th century and is the result of a recently finished FWF-funded research project “Exhibitions of Modern European Painting 1905-1915” by Rosenberg et al.. The covered period is considered as particularly interesting from an art historical perspective, as several avant-garde movements, such as Expressionism or Cubism, are represented in corresponding exhibitions. According to the project team, the data set can be useful for various research topics, including the study of artists’ networks in a geographical and temporal context, the study of certain movements (“-isms”) and their role in particular exhibitions, the study of exhibition strategies of painters and the study of interaction of art criticism with exhibitions [DoMa].

The data set is publicly accessible on the project website and is - apart from a simple map view of the exhibitions - mostly restricted to a query-based exploration. The set is very comprehensive and contains information about more than 200,000 works (catalogue entries), which were exhibited by more than 13,000 artists in more than 1,300 exhibitions, predominantly in the time period of 1905 - 1915. A smaller number of exhibitions in the years 1904 and 1916 are also included. Geographically, there is a strong focus on exhibitions in European cities, which make up a large part of the total of 100 cities in the data set. In addition, some exhibitions on the American continent and one exhibition in Japan are also part of the database.

This master thesis aims to focus on the study of artists’ social networks in a geographical as well as temporal context. To this end, we propose network science as well as visual analytics techniques for an in-depth analysis of the data set.

Network science, as described by Barabási [Bar16], is a scientific field which has gained enormous importance in recent times in order to better understand complex systems. The origins of graph theory, the mathematical foundation for network science, can be traced back to the 18th century.

Visual analytics, as defined by Keim et al. [KMSZ09], can be described as an interdisciplinary field aiming to combine automated data analysis techniques with human examination utilizing interactive visual representations of the data.

Using state-of-the-art network visualization techniques, visual analytics can facilitate the understanding of spatiotemporal social networks.

We build upon several publications which have already successfully incorporated network visualization techniques for art history research. Most of these publications limit themselves to a few tools to visualize networks. In particular, they focus entirely on basic node-link diagrams, a very common way to represent networks (see Chapter 2.2.6). With respect to the state of the art in the field of network visualizations, we believe that the precise geographic as well as temporal analysis of social networks in the art history domain can be improved by using a combination of more innovative interactive visualization methods. Thus, we propose the visual analytics application *Exhibitions Explorer*, which includes a node-link-diagram-based tool for the creation of distinctive, small-scale networks and a map view to display the locations of exhibitions. Our application allows art historians to analyze the self-created networks in two ways: Using a node-link diagram or a matrix-based representation. To this end, we introduce a novel matrix-based visualization approach, which is suitable for dynamic, undirected social networks with a low number of individuals. As art historians are often interested in the analysis of rather small networks consisting of less than 50 individuals [LABV18], we believe that this approach is particularly beneficial for art history and related domains. The two visualization approaches for the study of small-scale historic networks are evaluated and compared to each other using an insight-based approach as introduced by Saraiya et al. [SND05], which has a different focus than the commonly used task-based evaluations. *Insights*, as defined by Saraiya et al. [SND05], refer to single observations made by the participants. The characteristics of such observations can be quantified and encoded for evaluation purposes. The authors also draw a comparison to task-based evaluations, which primarily focus on measuring accuracy and efficiency in predefined benchmark tasks. The insight-based evaluation method is more qualitative in nature and aims to simulate real-world scenarios in which end users of an application explore a data set in an open-ended manner. We believe that this evaluation approach allows for a more meaningful evaluation with respect to the characteristics and requirements of art history research.

1.1 Research Question

The work aims to answer following research question:

What are appropriate ways to adopt visual analytics for exploration of networked data and gaining insights in the domain of art history research?

To answer our main research question, we further subdivide it into sub-questions, which

we aim to answer in the scope of this thesis.

- *Which state-of-the-art network visualization techniques or combination of such techniques are effective to help answering typical domain-relevant questions in art history research?*
- *To what extent can domain-relevant insights be gained through interactive visual data exploration?*
- *Are matrix-based visualization layouts a suitable alternative to node-link layouts for studying small historical networks?*

1.2 Expected Results and Contributions

The goal is to develop a prototype that utilizes different visualization techniques to support art historians in the best possible way for their research. Furthermore, we also aim to answer our research question by conducting the methodological steps described in the following Section. To the best of our knowledge, the main contributions in this thesis are:

- Discussing the current network visualization approaches in the art history domain with respect to the state of the art in network visualization and outline open topics for future research.
- Introducing a visual analytics framework to study social networks in the art history domain.
- Introducing a novel matrix-based visualization approach to study undirected dynamic networks.
- Evaluating the network visualizations using an insight-based approach, which is the first insight-based evaluation applied in the domain of art history.

1.3 Structure and Methodology

We introduce basic concepts in graph theory and network science in Chapter 2.

In Chapter 3, we discuss applications of network visualizations in the domain of art history research as well as the state-of-the-art visualization techniques for networks. To find literature for both topics, we carried out parts of a *systematic literature review*, following the guidelines by Okoli [Oko15].

We discuss the prototype design in Chapter 4. The prototype design is based on the findings of the literature review, which will be discussed with respect to the *design triangle* framework proposed by Miksch and Aigner [MA14].

The prototype implementation is discussed in Chapter 5.

In Chapter 6, we discuss all features of our prototype *Exhibitions Explorer*. In particular, we present the main components of the application and some exemplary workflows to demonstrate how the application can be used to gain domain-relevant insights.

In Chapter 7, we discuss the evaluation of the prototype using the insight-based evaluation methodology. We first introduce the methodology as proposed by Saraiya et al. [SND05] and its benefits compared to other evaluation approaches. Then, we describe the evaluation protocol and the results.

In Chapter 8, we summarize and discuss our findings and evaluation results with respect to the above-described research sub-questions.

In Chapter 9, we briefly discuss future work and conclude this thesis.

Finally, the Appendix contains research data collected during the evaluation sessions.

Basic Concepts in Graph Theory and Network Science

In this Chapter, we give a very brief introduction into the key concepts of graph theory. Then, we focus on network science and its terminologies and notations.

2.1 Foundations in Graph Theory

A *graph* as defined in graph theory¹ consists of a set of vertices V and a set of edges E . A commonly used notation is $G = (V, E)$.

If two vertices u and v are connected by an edge e , they are both incident to e and *adjacent* to each other. A graph G can also be represented by an *adjacency matrix* $A(G) = (a_{ij})$, where $a_{ij} = 1$, if v_i and v_j are adjacent to each other and $a_{ij} = 0$, if there is no edge connecting v_i and v_j [Mer00].

A graph $G = (V, E)$ can have an *orientation*, which is given by a function $f : E \rightarrow V \times V$, so that for each edge $e \in E$ connecting two nodes u and v with $u, v \in V$, $f(e)$ is either (u, v) or (v, u) . *Directed graphs* are graphs with a non-empty set of edges and a predefined orientation. *Oriented graphs* are directed graphs with the additional condition that two pairs of vertices can be connected by at most one oriented edge [Mer00].

A graph G is called a *complete graph*, if $E = V^{(2)}$. The complete graph of a graph with n vertices can be denoted by K_n [Mer00].

In computer science, *adjacency lists* are the most popular way to store graph data,

¹We will follow the definitions and notations as given in *Graph Theory* by Merris [Mer00].

as many algorithms can be implemented efficiently using this format. In case of directed graphs, the adjacency list of a graph G consists of four lists in total, including a list of n vertices, a list of m edges, a list of incoming edges $(u, v) \in E$ into all vertices $v \in V$ and a list of outgoing edges $(v, w) \in E$ starting from all vertices $v \in V$ [Val22].

2.2 Network Science

2.2.1 Terminologies and Notations in Network Science

Even though the terms *networks* and *graphs* are often used synonymously, Barabási [Bar16] points out that a distinction between these two concepts can indeed be made. Networks mostly denote “real” systems, such as social networks, while graphs usually refer to their mathematical description. Analogously, the terms *nodes* and *links* in network science correspond to *vertices* and *edges* in graph theory. Other synonyms for *nodes* include *agents*, *items* and *entities*, while *links* are often referred to as *relationships*, *ties* and *connections* [HSSH20].

Barabási [Bar16] also introduced N and L as important parameters in network science, which refer to the number of nodes $|V|$ and edges $|E|$ respectively.

Unless we explicitly refer to graph theory, we will mainly use the terms *networks*, *nodes* and *links* in the following chapters of this thesis.

2.2.2 Special Types of Networks

A basic distinction of networks can be made by looking at the number of types of nodes and links. We distinguish between *homogeneous* networks, which contain only one type of node and link, and *heterogeneous* networks, which can have multiple types of nodes and links. Real networks, for example in the biology domain, are often heterogeneous [GJFM18].

In graph theory, *dynamic graphs*, as defined by Beck et al. [BBDW17], refer to a sequence of (static) graphs $\Gamma := (G_1, G_2, \dots, G_n)$ with a static graph being defined as $G_i := (V_i, E_i)$ and the indices denoting time steps $\tau := (t_1, t_2, \dots, t_n)$. We refer to this definition when discussing *dynamic networks* in the next chapters.

In *weighted networks*, the links are assigned specific weights. Considering the adjacency matrix representation, links are denoted by their weights, i.e., $a_{ij} = w_{ij}$. Weighted networks play an important role in many real systems [Bar16].

Multivariate networks are a special case of networks, which contain additional data about the nodes and links. Such networks play an important role in research, as the analysis of real-world networks often goes beyond the analysis of mere structures [NMSL19]. More formally, a multivariate network $|N|$ is defined by its graph $|G|$ and a set of attributes $A = \{A_1, \dots, A_n\}$ associated with the nodes and/or the links. Accordingly,

the attribute values for a given node $|v|$ are defined as $a^v = (a_{v1}, \dots, a_{vn})$, an analogous definition can be made for links [KPW14].

In real-world networks, nodes are often associated with specific locations. Schöttler et al. [SYPB21] define *spatial networks* as networks where the “nodes are associated with inherent and semantically meaningful spatial positions”, while *geospatial networks* are considered as a special subset of spatial networks in which the positions of the nodes are associated with locations on a planet. Links can also be associated with geographic information.

In graph theory, *bipartite graphs* can be explained by vertex coloring. A graph $G = (V, E)$ is colored *properly*, if all adjacent vertices are assigned different colors. The *chromatic number* of the graph is denoted as $\chi(G)$ and refers to the smallest number of colors needed for a proper coloring of G . A *bipartite graph* is a graph with $\chi(G) \leq 2$. It follows that a graph G with $\chi(G) = 2$ can be divided into two disjoint sets of vertices with each set assigned one color [Mer00].

In network science, *affiliation networks*, which can be mathematically expressed as bipartite graphs, are frequently studied. They consist of *foci*², which refer to focal points of interactions and people [EK10]. Examples of foci include a shared music class or, in the context of art history, a shared exhibition.

We can use *projections* in order to obtain a new network consisting of only one of the two sets of nodes of the bipartite network. For example, we can obtain a projected network of people where the links are based on shared foci in the bipartite network [EK10].

Ego networks, also called *ego-centric networks*, are special types of networks, which focus on the connections of one particular node. This node is called *ego*, and it has connections to all other nodes in the network, which are generally called *alters*. In the simplest case, only the links between the ego and alters are included in the network. This is referred to as a *1.0-degree ego network*. In a *1.5-degree ego network*, the links between the alters are also contained. Finally, a *2.0-degree ego network* additionally contains all connections of the alters, regardless of whether these nodes have a link with the ego or not. In this case, the alters can be considered as additional egos. Ego networks are a well-known concept in social networks and used in social media platforms such as *Facebook* to find other users [Hog11].

2.2.3 Graph Drawing and Trails

To visualize networks as two-dimensional node-link diagrams, the nodes must be given coordinates. This is referred to as *graph-drawing* [AP13].

Researchers often make use of visualizations to study movement patterns. To this end, *trails* are important for visualization purposes. Lhuillier et al. [LHT17] define trails

²The term was introduced by Feld in 1981[Fel81].

as “oriented curves”, which, in most cases, display “the motion of shapes in space”. More formally, a single trail can be defined as $t_i \subset \mathbb{R}^d$, while a *trail-set* is given by $T = \{t_i\}$. There are many practical examples for the use of trails, including the visualization of migration patterns [LHT17] or air traffic [PHT15].

As Lhuillier et al. [LHT17] point out, there are differences between graphs as data sets and trail-sets. Most importantly, graphs are spatially embedded only after the graph drawing process, while trail-sets are already accordingly defined. Despite this distinction, we will also briefly discuss trail-sets in this thesis. First, because the study of migration patterns is an important research subject in the humanities. Second, the visualization techniques used to visualize trail-sets can also be adapted for graphs (see Chapter 3.3.4).

2.2.4 Important Structures in Networks

Communities are a fundamental concept in network science. Communities generally refer to structures in networks, where nodes are more likely to link to each other than to nodes from other communities [Bar16]. In social networks, for instance, communities might form due to individuals having shared interests [CA20]. More formally, communities have two main characteristics, according to Barabási [Bar16]: (i) communities are connected subgraphs, i.e., all nodes in a community are linked via paths; (ii) the nodes are locally densely interconnected, which implies that they are more likely to link to each other than to nodes of other communities. Identifying communities in networks is computationally difficult and multiple algorithms have been proposed for this task.

A well-known concept in sociology is that of *weak* and *strong* ties. In social networks, the strength of ties depends on various factors. For example, a high level of trust or intimacy can lead to stronger ties. However, in order to understand large-scale social networks, weak ties must also be studied. Weak ties are, for example, simple acquaintances. These weak ties serve as bridges between more tightly knit groups and therefore allow large-scale social networks to emerge in the first place [Gra73].

2.2.5 Centrality Measures

In network analysis, *centrality measures* play a significant role in evaluating the importance of individual nodes in the network. Various measures can be found in the literature, Gómez [Góm19] describes eight different ways to determine the centrality of a node. In the following we focus on four centrality measures as defined by Gómez, which are also mentioned several times in the literature on digital art history (see Chapter 3.2.1).

The number of neighbours denotes the degree in an undirected network. Using adjacency matrices, the degree k_i of a node can be given as $k_i = \sum_{j=1}^N a_{ij}$. In directed networks, we can distinguish between input and output degrees, depending on the orientation of the links. The input and output degree are given as $k_i^{in} = \sum_{j=1}^N a_{ij}$ and $k_i^{out} = \sum_{j=1}^N a_{ji}$ respectively. A commonly used centrality measure is the *degree centrality*, which is simply denoted by the degree of a node $C_i^{(deg)} = k_i$. Analogous definitions can

also be given for the input or output degree centrality.

The *eigenvector centrality* can be seen as a more sophisticated variation of the degree centrality, as it also considers the importance of the neighbours. The eigenvector centrality can be defined recursively as $\lambda C_i^{(eig)} = \sum_{j=1}^N a_{ji} C_j^{(eig)}$, with λ being the proportionality constant.

A *path* in graph theory refers to a sequence of edges connecting a sequence of vertices with each other. Accordingly, the *shortest path* between two vertices is the minimum path needed to connect these vertices.

The *betweenness centrality* in network analysis refers to the importance of a node in communication flow. It is denoted as $C_i^{(betw)} = \frac{1}{(N-1)(N-2)} \sum_{\substack{s,d=1 \\ s \neq d \neq i}}^N \frac{\sigma_{sd}(i)}{\sigma_{sd}}$, with σ_{sd} being the number of shortest paths between s and d and $\sigma_{sd}(i)$ being the number of shortest paths between these nodes which go through the node i .

Finally, the *closeness centrality* is based on the distances d between a given node i and the other nodes in the network. Mathematically, it can be denoted as $C_i^{(clos)} = \frac{1}{\sum_{j=1}^N d_{ij}}$ with d_{ij} being the distance, i.e., the number of links of the shortest path between nodes i and j .

2.2.6 Basic Visualization Layouts for Networks

An intuitive way to visualize networks is to use node-link diagrams [BBDW17] (sometimes referred to more broadly as node-link layouts [NMSL19]). Another common way is to use a tabular layout [NMSL19], which includes visualizations of the adjacency matrices. A simple way to represent an adjacency matrix is to use a Table structure with rows and columns corresponding to the nodes and colored or saturated cells indicating that the two nodes are adjacent [NMSL19, BBDW17]. A third, less common way to represent a network is an adjacency list-based visualization [BBDW17], showing the entire list of nodes and their adjacent nodes next to them.

Let G be a directed network, with:

$$A(G) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

Figure 2.1 shows three different visualization approaches for network G . Figure 2.1a shows a node-link diagram, Figure 2.1b displays an adjacency matrix-based representation and Figure 2.1c an adjacency list-based representation accordingly. Since the network is directed, it can be well observed that the adjacency matrix is not symmetric in this case.

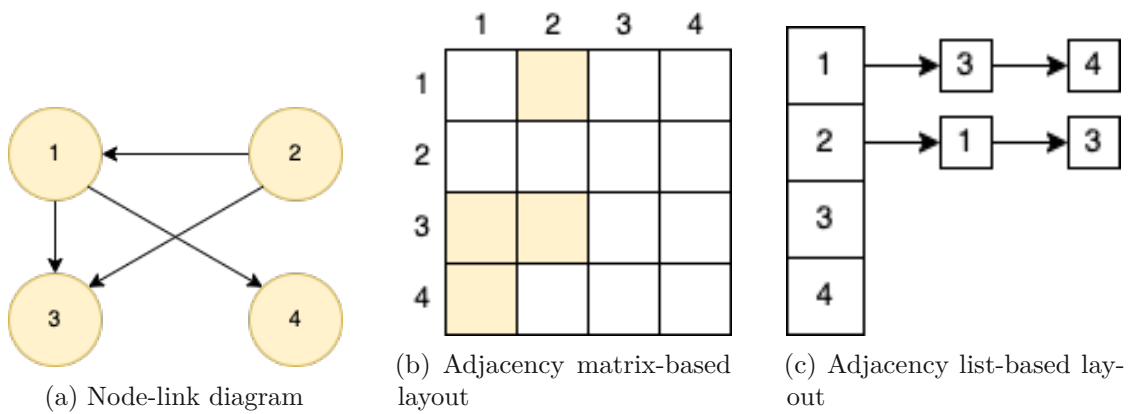


Figure 2.1: Three basic network visualization layouts.

Ghoniem et al. [GFC05] evaluated node-link diagrams in comparison to matrix-based visualizations in a controlled experiment. To this end, the participants had to perform seven generic tasks such as finding the node with the highest degree centrality. The findings revealed that node-link diagrams are particularly useful for small networks. Furthermore, node-link diagrams might also be beneficial for path-related tasks, e.g., finding a specific path in the network, especially in small or medium-sized networks. Matrix-based visualizations are more helpful for large and highly-connected networks.

Related Work

The use of information technologies in the humanities, including art history, has already been addressed in various works. These are often subsumed under the term *digital humanities*, a field which is already covered by some dedicated journals [Sve10].

In this context, *cultural analytics*, a term coined by Lev Manovich should also be mentioned. Manovich [Man16] described *cultural analytics* as the study of large data sets, utilizing both computational analysis and visualization techniques. The term, therefore, has much in common with *visual analytics*.

3.1 Research Methodology for the Literature Review

To identify related work, as well as, provide a theoretical foundation for the prototype, a systematic literature review (SLR) was partly conducted to gather relevant related work. In particular, we followed the guide provided by Okoli [Oko15]. His guide focuses on standalone SLRs in information systems research and provides a detailed overview of their major eight steps. Since a comprehensive SLR would be out of scope for this thesis, we formulated two research questions for the review and then mostly focused on the following three steps, which Okoli [Oko15] described in his guide:

- **Apply practical screen:** In this step, we defined inclusion criteria for the search process in order to reduce the number of reviewed articles to a manageable level, but also to ensure that the most relevant literature is included.
- **Search for literature:** In this step, we conducted the actual search process, applying the previously defined practical screen. We used three different online libraries for this purpose.

- **Extract data:** In this step, we extracted the data from the identified literature in order to answer the research questions.

We conducted the SLR to answer following two research questions in particular:

1. *How are network visualizations and visual analytics techniques for network exploration used in the domain of (digital) art history?*
2. *Which are the state-of-the-art visualization techniques for networks?*

3.1.1 RQ1: The State of the Art in Visualizations and Visual Analytics in Digital Art History Research

In order to answer the first research question, we decided to use the *ACM Digital Library*¹ and *ProQuest*² to find relevant literature. As ProQuest incorporates multiple databases, we decided to only browse the *Art, Design & Architecture Collection*, which includes multiple specialist indexes in the area of visual and applied arts. We considered following search query for both libraries:

("art history") AND ("visualization" OR "visual analytics" OR "network analysis")

We decided to use the keyword *network analysis*, as *network* on its own would have yielded too many results. We consider this to be justifiable, as all the keywords are linked by a logical *OR* operator. We only considered peer-reviewed literature, which mainly discusses the application of (network) visualization or visual analytics techniques for supporting researchers in the domain of art history. Excluding non-peer-reviewed, grey literature is not entirely risk-free, as *publication bias* must be taken into account. Publication bias refers to the fact that negative results are less likely to be published in peer-reviewed literature. It is particularly problematic in medical systematic reviews, as a large number of clinical trials are not published, resulting in a misleading impression of treatment effects [Pae17, RDP⁺13]. However, we estimate the risk of weighty publication bias to be low given the research question and the domain of the review. In addition, restricting the search to peer-reviewed literature in ProQuest allows for a significant reduction in the amount of literature found. As this work does not claim to apply a full-fledged systematic literature review, we believe that the restriction to peer-reviewed literature is justifiable. Furthermore, we only considered literature written in English and which we had institutional access to. Finally, we restricted the publication period to a 30-year window between April, 1992 and April, 2022.

The above-mentioned search query yielded a total of 652 results in ProQuest (applying the “peer-reviewed” filter offered by ProQuest) and 268 results in the ACM digital library when searched in “anywhere”. We reviewed the results by reading the title and abstract, only in some cases the full text was also considered. We identified a total of 13

¹<https://dl.acm.org/>. Accessed: 2022-11-27.

²<https://www.proquest.com/>. Accessed: 2022-11-27.

publications relevant for the above-stated research question. Most works were excluded as the filters were intentionally chosen to be very permissive and many results therefore did not have a sufficient relevance for our topic. One of these publications [BAA⁺19] is not discussed further, as it mainly refers to a visualization approach that is already covered in another publication. Due to its exceptionally high relevance for the research question, we decided to include one additional article [Kie17a] using a *backward search*, i.e., a work which was cited in one of the publications we found by searching the databases. We therefore ended up discussing a total of 13 publications for RQ1.

3.1.2 RQ2: State of the Art in Network Visualization

To answer the second research question, we used the *ACM Digital Library* and the *Wiley Online Library*. As Wiley's digital library covers literature on various subjects, we filtered for literature on the subject *COMPUTER SCIENCE*. We included existing surveys and/or reviews for network visualization techniques and the corresponding primary literature by performing a *backward search*. Once again, we only included peer-reviewed literature. We restricted the publication period to a 10-year window between April, 2012 and April, 2022, since a longer period would contradict our efforts to review the state-of-the-art visualization methods. We used following search query for both libraries:

("state-of-the-art" OR "survey" OR "review") AND ("network visualization" OR "graph visualization")

We found a total of 730 results in the ACM digital library and a total of 177 results in the Wiley Online Library, when searching in "anywhere". We identified a total of 10 publications relevant for our research question. Since some of these publications [WT17, SBMK14] have little relevance for historical networks in our view, they are not discussed further in this thesis. In addition, two publications [CLY17, HHS20] are not addressed further because they have substantial overlaps with other reports we included. Finally, a state-of-the-art report on multilayer network visualization [MGM⁺19] is not further addressed, as complex multilayered networks are not the focus of this thesis.

3.2 Network Analysis and Visualization in Digital Art History Research

3.2.1 Use of Network Analysis and Visualization Techniques

Social networks play an important role in art history. Lamqaddam et al. [LABV18] investigated to what extent familial and non-familial connections of historically significant individuals can be visualized. To this end, they developed a tool called *NAHR* (Networks in Art History Research), which builds on classic family tree visualizations. The visualization of time-related aspects turned out to be particularly challenging. To allow an interactive exploration through time, they decided to add a vertical slider. The users could select a specific year using the slider, allowing them to explore the professional and

god-parenthood connections in this year, in addition to the family links. The slider also highlights the number of events happening in a certain year by displaying one square for one event each in a bar plot. The links and squares are assigned the same colors.

Of particular note is that, while the database contains data on several thousand individuals, the visualization tool only examines a subset of 32 individuals. The authors based their decision mainly on a survey they conducted with 11 people in the art history domain, including professors and PhD students. For the overwhelming majority of users interviewed, it was sufficient to study smaller communities consisting of fewer than 50 individuals for their research activities.

Saint-Raymond and Courtin [SRC17] investigated networks of artists, buyers and sellers utilizing linked open data platforms and using the software package Gephi³ for visualization. They used a data set containing data on sales of modern paintings in Paris from 1850 to 1930. The records were extracted from auction catalogs. In order to enrich the original data with attributes such as artist's gender, nationality or date of death, they used the *Union List of Artist Names* and Wikidata as open data platforms. The authors described a total of five networks they created. Depending on the network, the artists, the buyers of the works or both together were visualized as nodes. The links were created on the basis of the auction data. For example, in two networks, a link between two artists implies that at least one of their works was sold in the same auction. Additional visualization elements were included to convey more information. For example, the size of a node depends on the average auction price achieved for the respective artist's works. The colour of a link darkens the more sold works two artists had in common. Time-related aspects were not considered by the authors, as they focused only on the 20 auction sales, which took place in 1868.

It should be noted that linked data technologies are generally becoming increasingly important in the digital humanities. In an article, Macken [Mac21] gives an overview of several current projects in the USA in which Wikidata in particular plays an important role. For instance, she mentions the *Wikidata Pilot Project of the Program for Cooperative Cataloguing (PCC)* as an example. Numerous libraries and organizations from the USA are participating in this large-scale project, where linked data is created on, among others, artists and exhibitions. Once the data is available on Wikidata, it can be used free of charge by a wide audience worldwide, for example for visualizations.

Leone and Vierthaler [LV20] used network analysis techniques to examine Pope Innocent X Pamphilj's role as a patron in the seventeenth century. In particular, they examined the network of various professionals, who were involved in significant construction projects under Innocent X's patronage. For the network creation, they extracted data from historical documents of fifteen projects, resulting in a network of 226 professionals. In addition to architects, painters, sculptors, and artisans, such as brick cutters, were also included in the network to enable a more comprehensive view of these building projects. For the links, they decided to make use of network projection.

³<https://gephi.org>. Accessed: 2022-11-27.

They first created a bimodal network consisting of the projects' work locations and the professionals being involved. The network resulting from the subsequent projection then contained only the professionals as nodes, whose links were based on contemporaneous presence in the respective locations. Closeness and betweenness centrality were then chosen as metrics to investigate the role of certain individuals in the network. The network was also visualized as a simple node-link diagram using Gephi, with the size of the nodes corresponding to their betweenness centrality. The results of the network analysis supported already existing hypotheses, but they also provided new insights. For example, the influence of Bernini, a designer, was probably underestimated in the preexisting literature, as he was not a favorite of the Pope, but nevertheless occupied an important position in the network. It is notable that in contrast to Saint-Raymond and Courtin [SRC17], Leone and Vierthaler also studied the yearly changes in the network.

Rice [Ric17] studied collaboration networks of artists for paintings, which were created in the Mughal Empire under the reign of ruler Akbar. Focusing on a few illustrated manuscripts, Rice identified different-sized teams consisting of designers and colorists, who worked on a varying number of paintings. The artists were represented as nodes in the multiple social networks she created. As in previous works discussed, the betweenness centrality in particular served as an important measure for the network analysis. Rice also mentioned the degree centrality as a potentially interesting measure, but ultimately decided to study the betweenness centralities of the nodes, as the measure does not entirely rely on the quantity of links, but also takes their importance into consideration. Furthermore, Rice also studied communities, which she referred to as "clusters". Finally, she addressed the concept of weak and strong ties. Weak ties, in this sense, were defined as collaborations, which took place rarely or even once. These weak ties allowed for a very diverse artistic collaboration, which ultimately also led to "stylistic fusion" [Ric17]. Rice visualized her networks using Gephi. In order to highlight important individuals, nodes of particularly high betweenness centrality were larger and more saturated in her visualizations.

*Palladio*⁴ represents a noteworthy visualization tool in the field of digital humanities. The Palladio app is an ongoing research project developed by Stanford University's *Humanities + Design Research Lab* [Hum]. The source code of the framework⁵ as well as the web application⁶ are available under the BSD 3-clause license on GitHub. The tool is described in detail in a journal article by Conroy [Con21]. In this article, Conroy distinguishes Palladio from other visualization tools such as Gephi, mentioned earlier. Palladio does not support any sophisticated network analysis techniques and focuses strongly on the visual representation of historical data. To this end, Palladio has powerful visualization capabilities and takes both temporal and spatial aspects of data into account. The tool essentially consists of four visualization components. In the first component,

⁴<https://hdlab.stanford.edu/palladio>. Accessed: 2022-11-27.

⁵<https://github.com/humanitiesplusdesign/palladio>. Accessed: 2022-11-27.

⁶<https://github.com/humanitiesplusdesign/palladio-app>. Accessed: 2022-11-27.

which is a map view, users can choose between a node and a node-link representation. The latter can be used, for example, to track movement patterns of important individuals on the map by linking place of birth to place of death. A notable feature of Palladio is the filter function, which also allows time filtering. Users can filter by time period or timeline, and thus visualize changes over the time. The timeline also acts as a bar plot and thus visually highlights years with particularly many events.

The second component consists of a graph view, which visualizes networked data as node-link diagrams. The diagrams are kept simple and, in contrast to Gephi, there is barely any support for analytical tasks and interactivity.

The table view consists of lists comparable to commonly used applications like MS Excel. This view is intended to provide a more in-depth analysis of the data and therefore complement the visualization tools discussed previously.

The last visualization component is the gallery, which is intended to display images related to the studied data. For example, researchers in the humanities might be interested to look at the portraits of important individuals.

According to Conroy, the combination of different visualization tools, as well as the spatial and temporal traceability of the data, make Palladio a very popular tool in the humanities. Nevertheless, Conroy also highlights some shortcomings. For example, data cannot be edited or cleaned in the tool. There are also comparatively few customization options. Conroy points out that Palladio has its strengths in data exploration, while deeper analysis may require more sophisticated visualizations as well as more interactivity.

McCabe [McC19] applied network analysis and visualization techniques to analyze Hans Rottenhammer's career network by studying his relationships with influential persons in the arts market in a comprehensive manner. As a first step, she examined Rottenhammer's patrons and anonymous purchasers in a geographic context. Using a simple node-link diagram created with the tool Palladio, she showed that Rottenhammer, both earlier and later in his career, continuously sent works to patrons in four European cities. As a result of this, she linked the two cities in which Rottenhammer worked with the hometowns of his purchasers. The corresponding visualization can be seen in Figure 3.1. In a further step, she studied Rottenhammer's social network, where she was particularly interested in the roles of the individuals for Rottenhammer's career as well as their personal backgrounds. To this end, she studied the betweenness centralities of the nodes in order to identify key intermediaries. She also briefly discussed a few communities in Rottenhammer's social network. Similarly to previously discussed works, McCabe visualized the social network using a node-link diagram, for which she used Gephi. Figure 3.2 shows a close-up of the social network, which displays Rottenhammer's relationship with Hieronymus Ott, whose broker role in the network results in a correspondingly high betweenness centrality.

Noble et al. [NVCD22] argued that social media analysis can also be conducted in an art historical context. People share their experiences in museums on social media, such as Twitter and Facebook, which can subsequently be analyzed for research purposes. For instance, it is possible to investigate how works of art are interpreted in the present

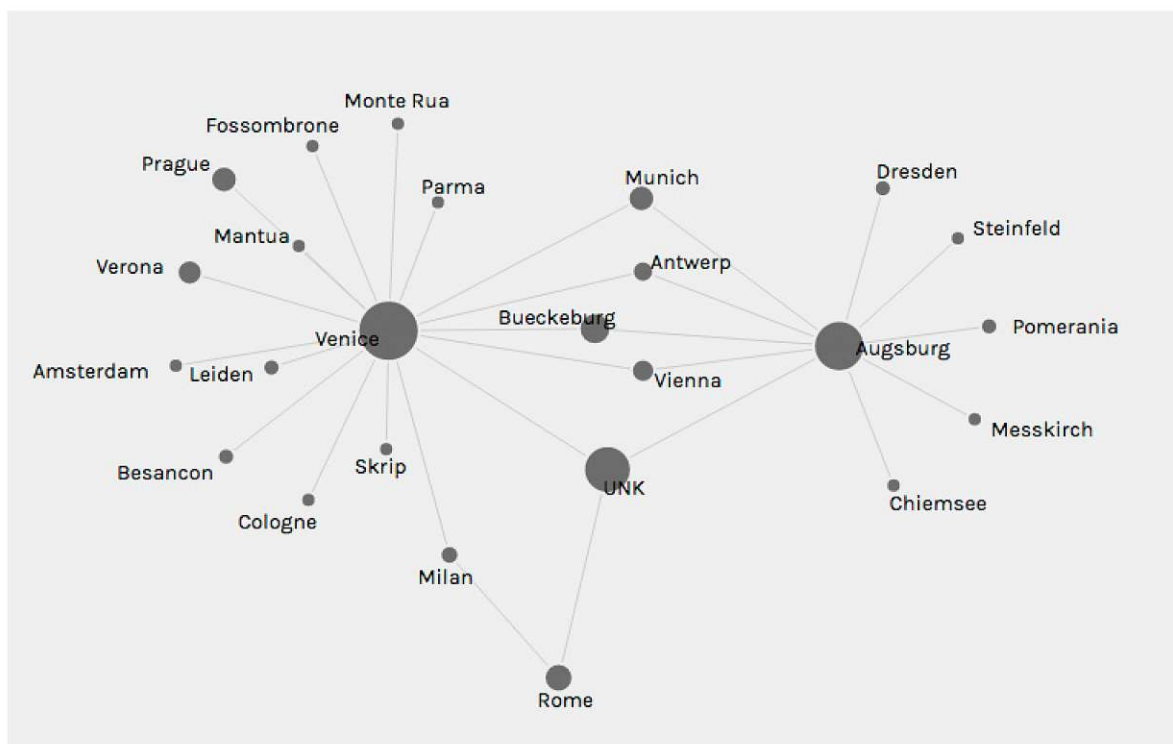


Figure 3.1: Rottenhamer worked in Augsburg and Venice. These nodes are associated with the hometowns of the patrons as well as unknown collectors (represented as “UNK”), who purchased Rottenhamer’s works. The visualization was made by McCabe with Palladio [McC19].

in comparison to the past. To demonstrate this, the authors constructed networks for which they extracted information from tweets. One of these networks examined the co-occurrences of keywords in tweets referred to *Rokeby Venus*, a famous work of art exhibited at the National Gallery in London. For this purpose, they used *Infranodus*, a text network analysis tool developed by Paranyushkin [Par19]. *Rokeby Venus* was damaged in 1914 by Mary Richardson, an activist in the British women’s suffrage movement. Noble et al.’s [NVCD22] network analysis showed that even in the present many museum visitors strongly associate the work with the suffrage movement. At the same time, the work seems to be strongly associated with feminism as well as female beauty. The authors used a node-link layout to visualize this network. They also highlighted a total of seven identified communities in color. However, these were not discussed in detail in the article.

Another application of network visualization in the art history domain can be the study of the historians themselves. For example, it is possible to construct a bipartite network based on conferences and their speakers. With the help of projection, a network

3. RELATED WORK

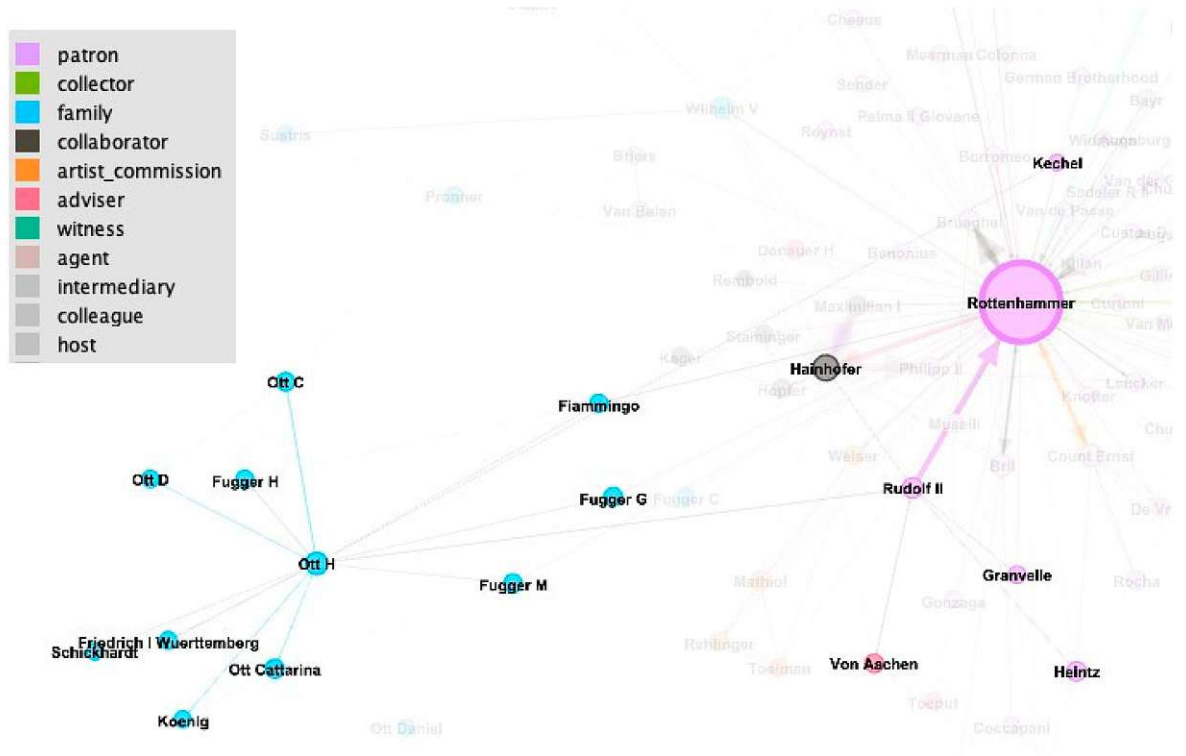


Figure 3.2: Close-up of Rottenhammer’s social network, focusing on his relationship with Hieronymus Ott [McC19].

of co-presenters can then be created, which may provide insights into the scientific domain. The construction of such a network was discussed by Burckhardt [Bur17] in an article. He worked with two data sets, including a set based on conference announcements by *H-ArtHist*, an internationally important “communication and information network for art historians”[Art] based in Germany. Burckhardt was able to extract more than 26,000 individuals from these announcements, with around 5,000 people who presented at least twice. However, he also emphasizes that this number should be considered with caution due to difficulties in the extraction process. Subsequently, the network was created, which initially exclusively contained persons appearing at least twice in the announcements. Among other findings, the analysis revealed that the nodes were mostly connected due to geographic reasons and not because the researchers worked in thematically overlapping areas. A node-link diagram was chosen for the visualization, with communities colored differently and the size of the nodes being proportional to their eigenvector centralities.

Networks including corresponding visualization techniques can also be used in the field of art education.

Grodoski et al. [GWG17], for example, studied publications in the journal *Studies in Art Education*, particularly because concerns arose that research in this area was isolated.

They proposed bibliographic coupling, which refers to two works citing a common third work [Kes63], and co-citation, which indicates how often two works are cited together in a third work [Sma73], to study past structural relationships in this journal. The subsequent visualizations of the created networks and subnetworks, via node-link-diagrams, revealed different citation structures. In particular, in the period from 1959 to 1968, many of the journal's publications were not part of a larger network, therefore implying a certain degree of research isolation in the early years of *Studies in Art Education*.

Network visualizations can also be used to teach art history. As Sweeny [Swe13] points out in an article, network visualizations have great potential for pedagogical purposes in the art education domain. In particular, techniques of network analysis can be used to facilitate the understanding of complex relationships between art and visual culture.

3.2.2 Problems and Challenges in the Visualization of Networks

In the previous Chapter, it was already mentioned that Lamqaddam et al. [LABV18] deliberately visualized only a small subset of the data they had access to, as the art historians they interviewed preferred it for their research. Saint-Raymond and Courtin [SRC17] also argue that large-scale visualized networks are both a curse and a blessing. A single visualization can display a massive amount of information to the observer. However, an increasing number of nodes, for example, makes the analysis of the data set more difficult. They suggest removing nodes with less relevance from the visualization. They also state that complexity becomes an even bigger problem once temporal changes over a longer period of time are taken into account. Despite the innovative nature of network visualization techniques, Saint-Raymond and Courtin highlight the continuing importance of traditional statistics and simple visualizations such as maps and graphs, which can complement networks very well.

A few works reflect on network analysis and visualization on a more critical basis, in particular when it comes to the underlying data used for such visualizations. As Porras [Por17] argues in an article in the *Artl@s Bulletin*, art historians must take care of biases when studying networks and their visualizations. Network visualizations rely on an underlying data set in the art history domain, which then again rely on limited written historical sources. Porras [Por17] specifically addresses how political, social, and cultural power structures can give a distorted picture of history. For example, for certain data sets, the power imbalances between colonizer and colonized need to be critically reflected upon. If networks are visualized without considering such aspects, there is a risk of biases becoming solidified.

Similarly, Kienle [Kie17a] states that data are not neutral, and we need to consider the underlying assumptions in according visualizations. Furthermore, she critically reflects on the focus on spatial aspects in network visualizations, which in turn neglects time-related aspects.

Kienle also discusses issues related to network visualizations in the art history domain in an interview with Johanna Drucker, who is a professor at the University of California, Los Angeles, and Miriam Posner, who is an assistant professor at the same university [Kie17b].

Druckner emphasizes that data in humanities very often require subjective interpretation, while information visualizations are often wrongly characterized as presentation of facts. Posner also highlights that a large amount of “context and perspective” are not taken into account in visualizations.

3.3 State-of-the-Art Network Visualization Techniques

3.3.1 Dynamic Network Visualization

Beck et al. [BBDW17] conducted a survey on dynamic graph visualization in 2016, where they also emphasized the importance of considering time-related aspects in visualizations, as networks are rarely static in the real world.

The authors define two main ways to visualize dynamic networks in their taxonomy, namely *animation* and *timeline* approaches. Furthermore, they also mentioned a few hybrid approaches which combine both animation- and static timeline-based representations, however we will not discuss these examples in detail.

Animation-based Approaches and the Mental Map

In animation-based approaches, the timestamps (see 2.2) of the dynamic network are mapped to the visualization times of the respective network. For interactive exploration, sliders might be used.

For node-link diagrams, Beck et al. [BBDW17] mention that the basic layout of the graph sequence is often kept the same during animation, so that users can conveniently track important changes. This is also referred to in the literature as *preserving the mental map*. The term *mental map* is used in various disciplines, including geography and psychology [Gra76]. Beck et al. [BBDW17] refer to the mental map as the “the abstract structural information a user forms by looking at the layout of a graph”. Based on a definition by Coleman and Parker [CP96], Archambault and Purchase [AP13] discussed mental map preservation as the effort to leave the location of nodes as well as links as unchanged as possible in dynamic graphs. They mentioned multiple experiments, which evaluated the importance of mental map preservation for different task types and concluded that there is not sufficient evidence to assume that it would actually help in the analysis of dynamic graph series. However, they also stated that a larger number of nodes and links than used in the discussed experiments would probably lead to a greater value of mental map preservation.

Beck et al. [BBDW17] mostly discuss animated node-link-diagrams in their survey, as adjacency matrices were rarely used in this context. However, they mention one conference paper by Rufiange and Melançon [RM14], who used an animated matrix approach to visualize design changes in software projects. In this paper, they illustrate two general approaches to facilitate the tracking of changes in dynamic graphs, using

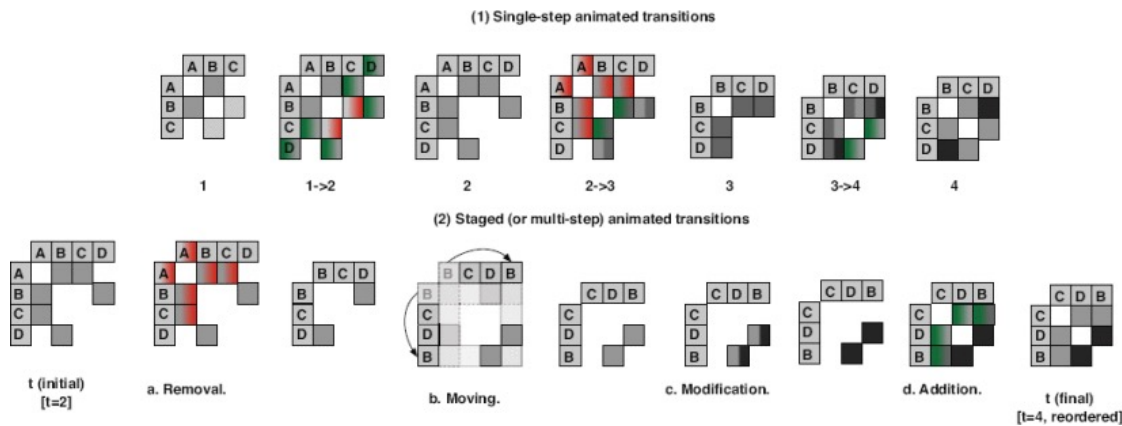


Figure 3.3: Single-step and multi-step animations to highlight changes in dynamic networks. [RM14] © 2014 IEEE

a matrix-based representation. This can be seen in Figure 3.3. The first approach is a single-step animation, where all the changes are shown at once. Figure 3.3 illustrates three of these animations. New links are introduced with a green color, while disappearing links are initially highlighted with a red color. The second approach consists of multiple steps, which allows to track changes more accurately. In the first step, the links disappear, again introduced by marking the corresponding cells red. In the second step, the position of some nodes might change in the matrix representation. The third step highlights changes of existing links, e.g., the links might become more saturated due to an increase of the link weight in case of weighted networks. Finally, new links are added, introduced by a green highlighting of the respective cells.

Timeline-based Approaches

Another approach is to display the sequence of dynamic graphs on a timeline. As Beck et al. [BBDW17] point out, this is mainly associated with one disadvantage, namely that plotting multiple graphs requires corresponding space.

With respect to node-link diagrams, the authors list the juxtaposition of the diagrams, the stacking of the diagrams on top of each other, and the integration of the timeline into the diagram as possible visualization options. Furthermore, adjacency matrices or even adjacency lists can be used for timeline-based representations. The latter is particularly interesting for this representation, as it does not take a lot of space to visualize networks as adjacency lists. The authors mention an article by Hlawatsch et al. [HBW14], who proposed an adjacency list-based visualization for dynamic weighted networks. One of the two layouts they proposed can be seen in Figure 3.5. The list of nodes for this network can be seen in the center. The timeline left from the list shows the incoming nodes in four time steps, while the timeline on the right is displaying the outgoing nodes, respectively. This approach can also be adapted for weighted networks, e.g., by adjusting the heights of

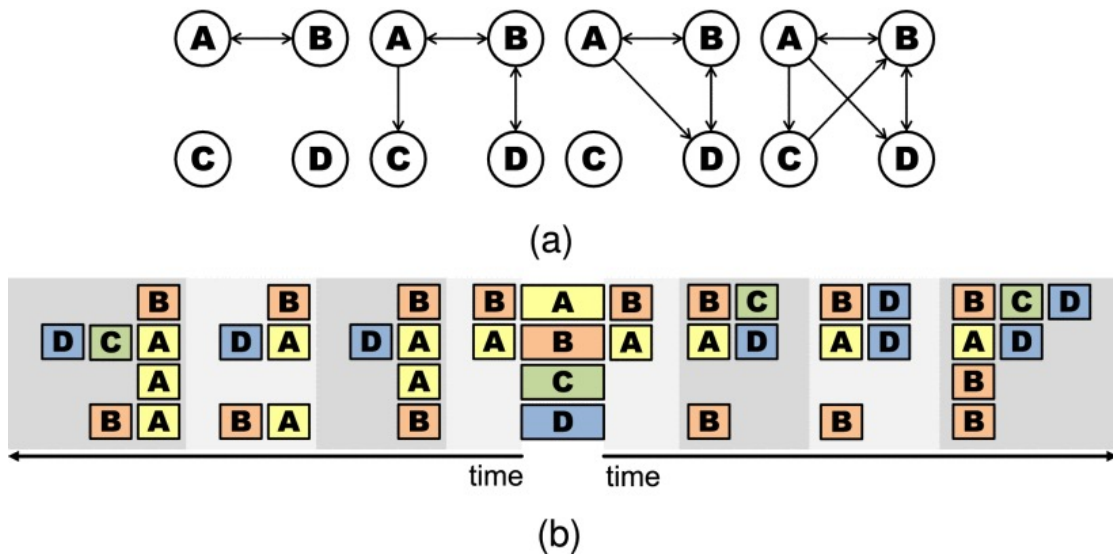


Figure 3.4: Timeline-based visualization using an adjacency list [HBW14]. © 2014 IEEE

the node cells according to their weight. In addition to the illustrated advantage of saving space, Hlawatsch et al. [HBW14] also showed experimentally that these representations can be equally effective or even superior to matrix-based representations or node-link diagrams for various tasks. For instance, the list-based visualization turned out to be more suitable to detect the distribution of outgoing and ingoing links over the nodes than the other two mentioned visualization approaches. However, they also pointed out that the list-based representation has its limitations with respect to path-related tasks or the detection of communities.

3.3.2 Geospatial Network Visualization

In a survey, Schöttler et al. [SYPB21] summarized different ways to visualize geospatial networks, also considering interaction techniques for further exploration of the data. Their survey studies the visualization of geospatial networks in four dimensions.

The first dimension refers to the visualization of geographic information. The authors distinguish between two main categories of visual encodings, namely explicit and abstract encodings. Explicit encodings make use of maps, including both two- and three-dimensional representations. Abstract encodings, on the other hand, represent geographic information in a different way. The authors mention an article by Hennemann [Hen13], who discussed the advantages of using circular layouts to visualize complex geographical networks. Henneman demonstrated his approach by visualizing a network of cities, based on office networks of globalized firms. The circular layout can be seen in Figure 3.5. The nodes are grouped by approximate geographic location. A special feature of this layout is the comparatively simple representation of further, non-geographical

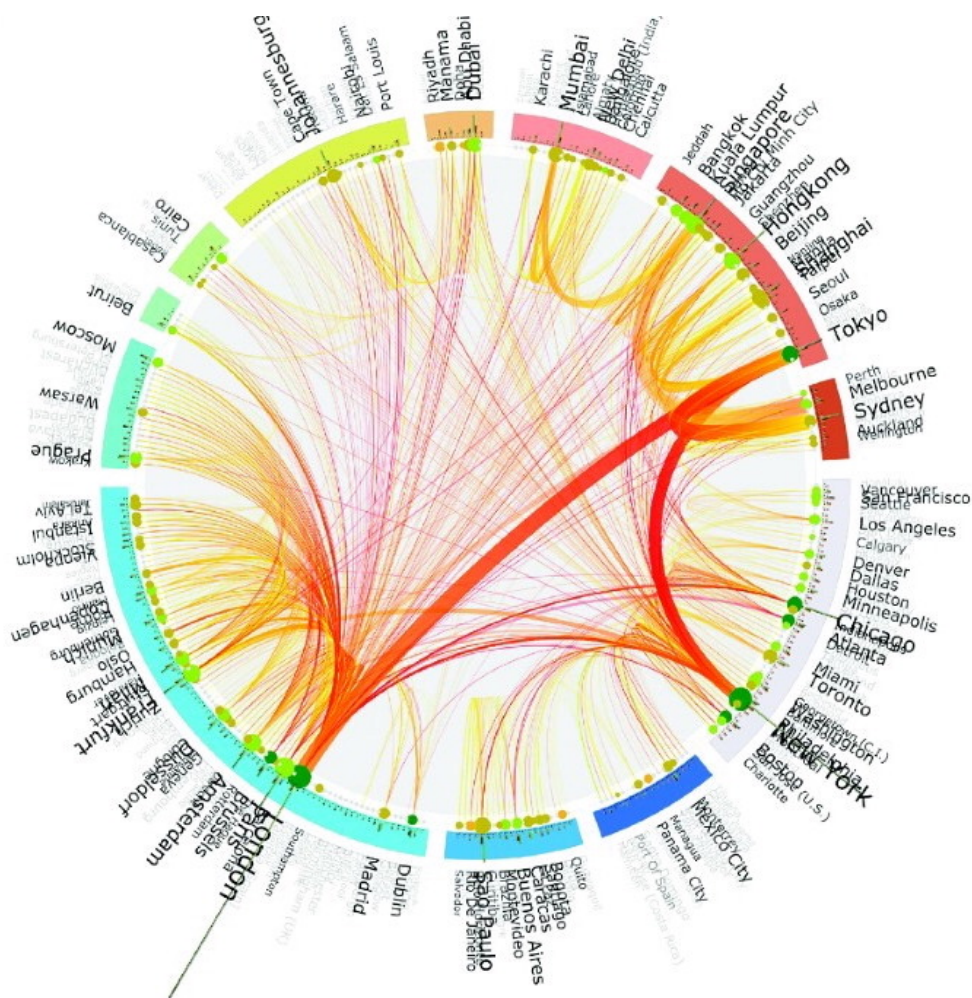


Figure 3.5: A circular layout to visualize a network of cities [Hen13]. Modified.

information. For example, the outer circle contains bar charts referring to ingoing and outgoing connections of each city. City names are sized according to their degree centrality, while their opacity is determined by the betweenness centrality. The colors of the links refer to the hierarchical position of the linked offices. Hennemann states that the main disadvantage of this visualization type is the high effort required to create it.

The second dimension Schöttler et al. [SYPB21] described refers to the representation of the nodes and links of the network. They once again distinguish between explicit and abstract visual encodings for both nodes and links. The explicit encoding represents topological information using a simple mapping, typically displaying nodes and edges as single or aggregated elements in the visualization. Abstract encodings use special mappings, which are often associated with the aim of using space more sparingly.

The third dimension they discussed refers to the integration of the geographic and the network representations into one visualization.

Finally, the last dimension refers to interactivity. The authors particularly investigated the use of interactivity for exploration purposes. While they pointed out that non-interactive visualization techniques have certain benefits, as they are less resource-intensive and can be used in static media as well, interaction techniques are a good way to reduce high complexity in a visualization, which is why some techniques they discussed *need* interaction to display the data set in its entirety. An example includes the *in situ visualization* approach introduced by Hadlak et al. [HSS11]. They proposed the integration of multiple visualization techniques in an interactive manner in order to study large-scale dynamic networks. They demonstrated this approach on a mesh network, which consists of network devices, represented as nodes, and connections between them, which are represented as links. The corresponding visualization is shown in Figure 3.6. The geospatial visualization shows an overview of the mesh network in the German city Rostock, while the node-link and matrix-based diagrams offer a more detailed view of the corresponding subnetworks. Thus, the data set in its entirety is revealed only through interaction with the basic map visualization.

3.3.3 Multivariate Network Visualization

In a state-of-the-art report, Nobre et al. [NMSL19] discussed various techniques to visualize multivariate networks. They proposed a typology consisting of three core layouts and three operations applicable to these layouts to categorize the techniques. The layouts they discussed include node-link, tabular (including matrix-based visualization techniques), and implicit tree layouts. To further facilitate the analysis of the data set, operations can be applied on the data or the layouts. Additionally, multiple visualizations can be combined using different views.

A notable visualization tool the authors mention for the domain of social network analysis includes *Vizter*, which was proposed by Heer and Boyd [HB05]. They use a node-link diagram for visualization, relying on a force-directed layout. This layout visualizes networks in an aesthetically appealing way. Intuitively, a node in this layout is pulling other surrounding nodes, while the links can be considered as springs. This leads to a structure where communities are implicitly visualized, as highly connected structures move closer together. Furthermore, Heer and Boyd used on-node encoding to visualize the attributes of nodes. For instance, the sex of the persons was highlighted by coloring the corresponding nodes red or blue.

Matrix-based visualization techniques can be beneficial when visualizing link attributes such as weights. Alper et al. [ABHR⁺13] proposed a technique which allowed the comparison of two weighted networks in a single matrix. While the absolute edge weight is represented by color brightness, the comparison of edge weights is enabled by color

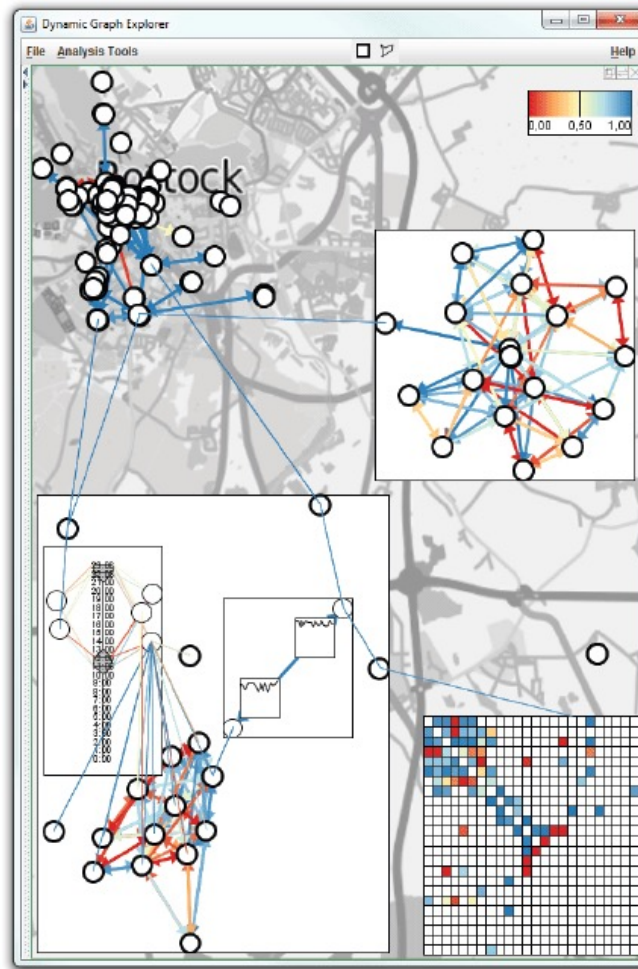


Figure 3.6: Integrated visualization approach of a mesh network in Rostock, Germany [HSS11]. © 2011 IEEE

differences of the inner and outer region of a cell. By conducting a controlled experiment, the authors showed that their matrix-based visualization approach was superior to node-link diagrams for certain comparison tasks.

NodeTrix, which was introduced by Henry et al. [HFM07], is a hybrid visualization technique, aiming to combine the benefits of node-link diagrams and matrices. Their approach makes it possible to visualize larger social networks in their entirety while allowing detailed analysis of community structures simultaneously. To this end, the overall topology of the network is mostly represented by nodes and links, while matrices are used to display communities. The general layout of this technique can be seen in Figure 3.7. Figure 3.7a - 3.7c show different approaches the authors considered to visualize links for the node-link component. They ultimately decided to display the single links and to not aggregate them, as shown in 3.7c. The users however can change the

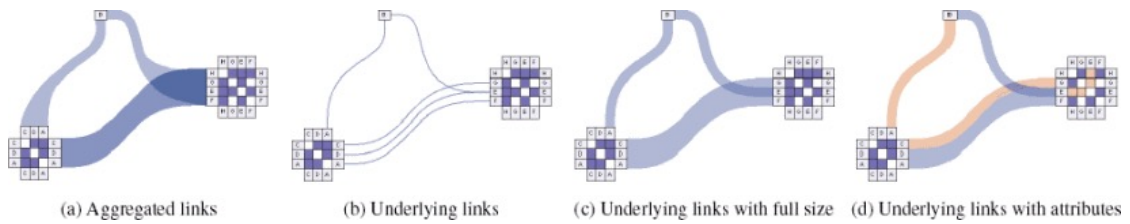


Figure 3.7: *NodeTrix*, a hybrid visualization approach [HFM07]. © 2007 IEEE

thickness of the links, with which a bundled-looking representation can be achieved. As shown in Figure 3.7d, link attributes can also be highlighted in *NodeTrix* by coloring the cells in the matrices and links in the node-link diagram.

Implicit tree layouts are used to visualize hierarchies, which can be represented very well as mathematical trees. To this end, multiple layouts have been proposed, while all of them encode links implicitly by aligning the nodes a certain way [NMSL19]. For circular layouts, the links are displayed implicitly by aligning the child nodes exactly under the parent node [AH98]. While node attributes can be displayed very well using implicit tree layouts, e.g., by using different node sizes and colors, link attributes cannot be encoded due to their implicit representation [NMSL19].

3.3.4 Visualizing Special Structures in Network

Groups

Networks often consist of group structures, whose study can provide further interesting insights. In a survey, Vehlow et al. [VBW17] discussed multiple visualization techniques to highlight such group structures in networks. Group structures can be derived by the topology of the network or by attributes of the nodes, with communities being topology-based group structures. The authors discussed visualization techniques for both node groups and link groups. Furthermore, they distinguish between *implicit* encodings, i.e., methods which indicate group structures by aligning alike objects near each other, and *explicit* encodings, where group structures are highlighted by accompanying visualizations.

Node groups are frequently highlighted by altering the representation of single nodes. To this end, coloring is commonly used. Vehlow et al. [VBW17] found a total of 11 publications which relied entirely on coloring for visualizing group structures. Other techniques include juxtaposed visualizations, where the graph and a group are positioned next to each other, and superimposed visualizations using overlays and embedded representations, where the nodes of a group are aggregated into a single node.

According to Vehlow et al. [VBW17], link groups are less common than groups based on nodes. Nevertheless, there is a large number of visualization techniques which can be applied to relationship groups. A notable implicit technique is to order the columns and

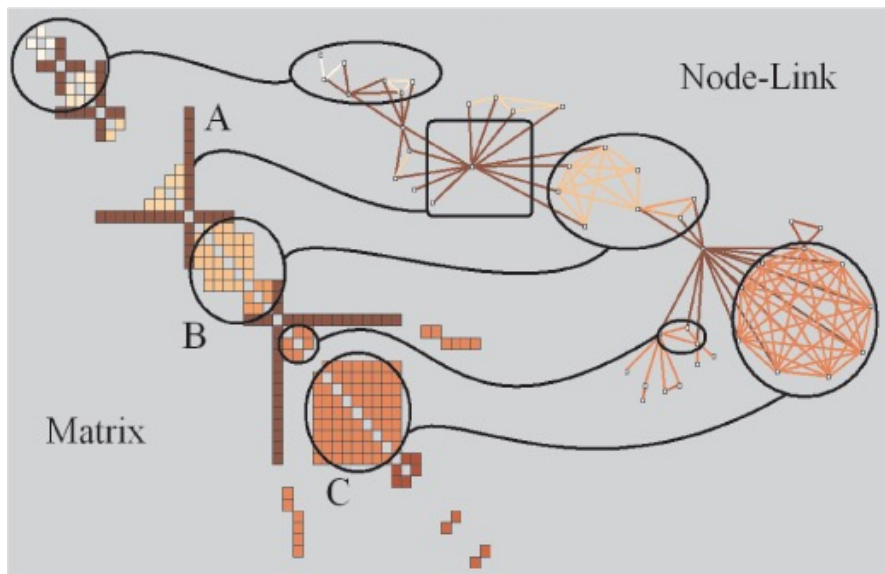


Figure 3.8: When coloring the cells in the matrix-based representation, the colors of the links in the node-link diagram change accordingly [HF06]. © 2006 IEEE

rows (representing the nodes) in an adjacency matrix in a way that link group structures are revealed. Multiple algorithms have been proposed to efficiently order the nodes for visual analysis in the best possible way [MML07], however it must be noted that the vast majority of graph layout problems are NP-complete when formulated as decision problems [DPS02]. As Vehlow et al. [VBW17] point out, the implicit representation of link groups is often combined with explicit encodings.

A notable example mentioned by them includes *MatrixExplorer*, a network visualization system proposed by Henry and Fekete [HF06], which makes use of both matrix representations and node-link diagrams. The system was developed to facilitate the analysis of social networks, taking into account the needs of researchers in various fields, including history. In particular, they also considered the benefits of matrices for community detection. To this end, they developed *MatrixExplorer*, which can represent networks as node-link diagrams as well as matrices. These representations are synchronized to fully utilize both representations. In order to identify communities, the users can automatically order the matrix representation of the network and thereby detect group structures. They can then color the cells that represent the links. When they switch to the node-link layout afterwards, the links are colored accordingly. An illustration of both representations can be seen in Figure 3.8.

Link and Trail Bundling

Especially for larger networks, it may be necessary to bundle links to make the visualization clearer and to facilitate the analysis of the basic structures. Lhuillier et al. [LHT17] conducted a survey on state-of-the-art bundling techniques for links (which they refer to as

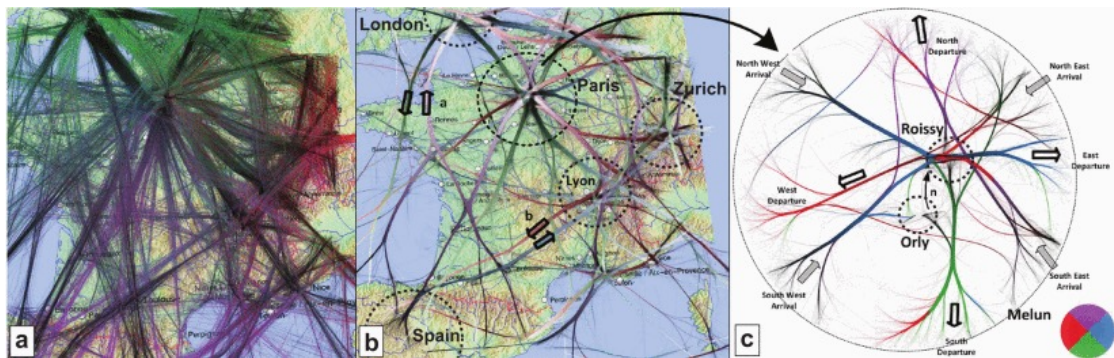


Figure 3.9: Attribute-Driven Edge Bundling applied on aircraft trails over France [PHT15]. © 2015 IEEE

links) and trails. In their taxonomy, they distinguish between two main bundling methods, namely *graph bundling methods* and *trail-set bundling methods*. However, the authors state that trail bundling methods can mostly be applied to generic graph drawings as well.

Attribute-Driven Edge Bundling (ADEB) is a notable bundling technique which Lhuillier et al. [LHT17] mentioned in their survey. It was proposed by Peysakhovich et al. [PHT15], who build upon the *kernel density estimation edge-bundling (KDEEB)* technique proposed by Hurter et al. [HET12]. For the bundling, the ADEB technique takes numerical attributes of the links, such as time or direction, into account and also supports bundling based on multiple attributes. The method is therefore suitable for the analysis of trail-sets, since trails can be associated with above-mentioned attributes. Peysakhovich et al. [PHT15] applied their technique on an aircraft data set, which contains recorded trails over France for an entire day. The ADEB method allows to include the flight directions in the visualization. The approach can be seen in Figure 3.9. Figure 3.9a shows the unbundled trails, which are already colored according to their direction. The darker colors show the start points, and the lighter colors show the end points. The bundled trails can be seen in Figure 3.9b. Thicker strokes indicate that a large number of trails are bundled in that area. The directional bundling makes it possible to track flights between major European cities such as London and Paris more clearly. Finally, a more detailed analysis of the flights over the metropolitan area of Paris can be seen in Figure 3.9c, which can be obtained by zooming into the area.

3.4 Discussion

3.4.1 Current Approaches in the Art History Domain

Network analysis techniques can be used in various ways for art history research. Most publications [LABV18, Ric17, SRC17, LV20, McC19, Bur17] we discussed focused predominantly on social network analysis. However, text network analysis [NVCD22] or the study of co-citation networks [Kes63] were also successfully used for research

purposes.

To obtain social networks mostly techniques of network projection [SRC17, LV20, Ric17, Bur17] were applied. Centrality measures played a fundamental role for the analysis of the respective networks. Betweenness centrality was the most popular centrality measure and was studied in five publications [SRC17, LV20, Ric17, McC19, Bur17], while degree centrality was studied in one [Bur17], eigenvector centrality in one [Bur17] and closeness centrality [LV20] in one publication, respectively. The study of communities also played an important role in some publications we found [Ric17, McC19, NVCD22, Bur17]. Finally, one publication [Ric17] briefly discussed the importance of strong and weak ties with respect to artistic collaborations.

Thematically, the areas of application range from the analysis of the relationships of historically significant personalities to the study of the art historians themselves. Specifically, teaching art history also represents an important use case for network visualization and analysis. In future, the importance of network analysis in the domain of art history can further increase, especially due to *Linked Open Data* technologies, which have already been successfully used for network analysis [SRC17] and play a significant role in current domain-relevant projects [Mac21].

All the approaches we discussed in Chapter 3.2.1 rely on node-link diagrams for the visualization of networks. Gephi was used particularly frequently [SRC17, LV20, McC19, Bur17, Ric17]. Palladio is an application that was developed specifically for the humanities domain and is also frequently used [SRC17, Con21]. Palladio supports geographic analysis of the data thanks to the map view. Furthermore, it supports time-based filtering for more accurate analysis of time-related characteristics. The network visualization, however, is kept very simple. Gephi, on the other hand, allows the use of network analysis techniques, such as the calculation and subsequent visual encoding of centrality measures.

3.4.2 Suitable Visualization Layouts for the Art History Domain

While node-link diagrams are generally said to be the most common visual representation of networks [NMSL19], several publications outline the potential benefits other visualization layouts, including matrices [HFM07, HF06, ABHR⁺13, HSS11], adjacency lists [Val22], tree layouts [AH98], and other layouts such as circular/radial ones [Hen13]. The combined use of node-link diagrams and matrix-based visualization layouts in interactive visualization applications [HSS11, HFM07, HF06] turned out to be particularly popular.

Considering the requirements in the art history domain, we think that the choice of a visualization approach largely depends on which type of networks is visualized. Homogeneous social networks can be visualized very well with node-link diagrams as well as matrices. Node-link diagrams were successfully used in numerous art historical publications we discussed. Matrices are beneficial for denser graphs [GFC05], which is specifically exploited in some applications like NodeTrix [HFM07] to better represent community structures. Node-link diagrams were also successfully used in the citation networks [Sma73] or text analysis networks [NVCD22] we discussed.

Another use case in the art historical domain is the visualization of movement patterns. This is also supported by Palladio, where the place of birth of a person can be linked to the place of death. In this case, maps and visualizing movements as links or trails can be considered as a common and popular layout. Sophisticated bundling techniques such as ADEB [PHT15] can help to visualize a large amount of trails. Different approaches include visualizations with the help of adjacency lists, which were also used by Hlawatsch et al. [HBW14] for the study of migration patterns, or more unique layouts such as circular/radial ones. The latter can be tedious to create, as they require a significant amount of effort for creation [Hen13].

Finally, tree layouts can be useful to visualize hierarchies, e.g., in visualizing family trees of historically significant persons. A similar use case was also described by Lamqaddam et al. [LABV18], but they decided to include extra-familiar links as well.

3.4.3 Techniques for Visualizing Specific Types of Networks

A closer look at state-of-the-art network visualization techniques reveals that a large variety of techniques can be applied for networks in the art history domain. The visualization of group structures [VBW17], for example, is of great importance for the analysis of social network communities. Link and trail bundling techniques [LHT17] can be significant in the study of migration patterns, for example. We consider dynamic [BBDW17], geospatial [SYPB21], and multivariate network visualization techniques [NMSL19] to be particularly suitable for the art history domain. Naturally, art historians are interested in when and where something happened. The necessity of these visualization techniques also becomes clear when looking at the art historical publications we examined. Representation of time-based data in network visualizations was highlighted as particularly difficult in two art historical publications we found [Kie17a, SRC17]. Simple maps were used in addition to network visualizations in two publications [LV20, McC19] we found, while one of them [McC19] additionally used a simple node-link diagram to visualize geospatial information. Palladio also incorporates a map view to allow the study of geographic data. Finally, basic multivariate network visualization techniques, such as node encoding, were already used in some art historical publications we discussed, as seen in Figure 3.2.

It should be mentioned in conclusion that the surveys we considered focus on different aspects or facets of networks, for example, the spatial nature or the temporal sequence of networks. However, the visualization approaches which are used to display different aspects are often identical. Multivariate network visualization techniques for instance can be used for geospatial networks as well. Which aspect is visualized in which form depends decisively on the network and on the underlying data model. For our prototype, we will describe the choice of visualization approaches with respect to the underlying data model in Chapter 6.

Prototype Design

The prototype design was developed according to the principles of the *design triangle* framework, which was proposed by Miksch and Aigner [MA14]. This high-level framework can help to identify the requirements of visual analytics solutions by considering three fundamental aspects:

- The specifics and characteristics of the data should be identified and considered.
- The general tasks that users complete in the course of their analysis should be understood.
- The needs of end users must be identified and taken into account.

These three aspects also result in three essential metrics for determining the quality of the visualization used, which Miksch and Aigner discussed with respect to the works of Mackinlay [Mac86] and Van Wijk [vW06]. *Expressiveness* refers to the relationship between data and the performed tasks. It is crucial that only what actually emerges from the data is visualized. The quality with which the data are presented to the user is measured by *effectiveness*. Finally, *appropriateness* evaluates the usefulness of a visualization for the completion of tasks in terms of efficiency.

Exhibitions Explorer visualizes a social network of artists on the basis of shared exhibitions. In the following four sections, we describe the fundamental design decisions of our application based on the characteristics of the design triangle. The choice of an adequate data model as well as further implementation details are described in Chapter 5. Finally, we will cover the resulting visualization design in detail in Chapter 6.

4.1 Data

Table 4.1: Three key entities and their properties.

Property	Classification	Entity	Description
Name	Nominal	Artist	Refers to the full name of an artist.
Nationality	Nominal	Artist	The data set contains the primary nationality of an artist. This information is important to examine the relation of the artist's background with the exhibition locations.
Sex	Nominal	Artist	Sex is included in the data set for most artists. One focus of research is the study of female artists in this period, as they are outnumbered by male artists in the data set.
Occupation	Nominal	Artist	Occupation denotes a very brief textual description of the artist. Based on common characteristics, such as artistic focus, interesting conclusions can also be drawn.
Title	Nominal	Exhibition	Refers to the title of the exhibition.
Start Date	Interval	Exhibition	Refers to the start date of the exhibition.
End Date	Interval	Exhibition	Refers to the end date of the exhibition.
Start Year	Interval	Exhibition	For the network visualization, the start year of an exhibition is essential, as this allows the creation of a dynamic network as a sequence of networks from 1904 to 1916.
End Year	Interval	Exhibition	The end year of an exhibition should be displayed to the user. However, for the network visualization, it makes more sense to create the sequence of yearly networks only based on the start year, as the networks can then be compared to each other in a meaningful way.
Number of Artists	Ratio	Exhibition	Indicates the number of artists who were represented in an exhibition.

Number of Catalogue Entries	Ratio	Exhibition	Indicates the number of catalogue entries for a single exhibition.
City	Nominal	Location	Refers to the exhibition city.
Country	Nominal	Location	Refers to the exhibition country. It makes sense to consider the countries additionally, as it allows to study the locations in a more broad manner.
Longitude	Interval	Location	Longitude of the exhibiting city.
Latitude	Interval	Location	Latitude of the exhibiting city.

The data set was provided to us by the DoME team [DoMb]. It contains information on modernist exhibitions in the years between 1904 and 1916, with a focus on exhibitions between 1905 and 1915. A major challenge was to reduce the large size of the data set to a small number of entities without losing any essential information, especially with respect to geography and time. In a first step, we therefore identified the key entities and their properties needed for the application. We classified the properties using the very commonly used “scales of measurement” typology proposed by Stevens [Ste46]. Afterwards, we classified these data using the proposed classifications for time-oriented data described by Miksch and Aigner [MA14], as it is crucial to appropriately visualize the network with respect to the given time period. Finally, the resulting network can also be classified by using the terms we introduced in Chapter 2.2.2, which is particularly important as we discussed the state-of-the-art network visualization techniques with respect to these terms.

4.1.1 Entities and their Properties

The DoME data set is very comprehensive and contains, besides data on exhibitions and artists, information about the exhibition catalogue entries, organizing bodies and venues. However, only a small subset of these data we identified are needed to visualize artists’ networks based on shared exhibitions. We therefore decided to include only three different entities in the application, namely *artists*, *exhibitions*, and *locations*. From our point of view, all three entities can be understood as nodes. The decision to model the entire data set as a graph will be explained and justified in detail in Chapter 5.

We then proceeded to identify the key properties of these entities. As mentioned before, we classified these properties based on a typology by Stevens [Ste46]. He proposed a classification of data into four fundamental “scales of measurement”, namely *nominal*, *ordinal*, *interval*, and *ratio*. *Nominal* denotes variables that can be named, i.e., categorized. *Ordinal* data can be categorized *and* ordered. *Interval* refers to ordered variables, where the order denotes proportioned distances (“equality of intervals” [Ste46]). Lastly, *ratio* denotes data that, additionally to the previously mentioned characteristics,

also have a true natural zero point.

The final set of properties we included in *Exhibitions Explorer* can be seen in Table 4.1.

4.1.2 Classification According to Characteristics of Time-Oriented Data

Another classification of the data, which Miksch and Aigner [MA14] described for the design triangle, can be done based on the characteristics of time-oriented data, as described by Aigner et al. [AMST11]. The authors defined four major characteristics of data based on the pyramid framework proposed by Mennis et al. [MPQ00]. The first characteristic refers to the *scale* of the data. In particular, the authors distinguish between *qualitative* and *quantitative* data. The *frame of reference* denotes whether the data has a *spatial* reference or is *abstract*. The *kind of data* denotes whether one is interested in the time points of state changes (*events*) or in the phases between them (*states*). Finally, based on the number of variables, one can distinguish between *univariate* and *multivariate* data.

Aspects, elements, and the hierarchical organization of time are discussed separately by Aigner et al. [AMST11], as time represents a crucial data dimension. Two of these time-related characteristics are considered by Miksch and Aigner [MA14] for the Design Triangle. First, *time primitives* denote *how* the data relates to time. *Instants* refer to individual time points in the data, while intervals refer to time ranges which are limited by instants. Finally, *spans* refer to arbitrary time spans in the data, e.g., a duration of 12 days.

The second characteristic refers to the *arrangement* of time. With respect to the work of Frank [Fra98], Aigner et al. [AMST11] distinguish between *linear* and *cyclic* time arrangement. Linear refers to a temporal sequence that has a beginning and an end, while cyclic refers to recurring periods, such as seasons.

Using this classification, we note that our application consists of both quantitative and qualitative, spatial, and multivariate data. Furthermore, we visualize events, as we are interested in state changes and not in the phases between them. Regarding the characteristics time, our application visualizes the network in a *linear* manner, i.e., from 1904 to 1916. Furthermore, we visualize *instants*, i.e., the network at one point (=year) in time. It would also be possible to take the exact exhibition duration into account and therefore visualize intervals instead of instants. However, for many exhibitions, the data set does not include exact information about the exhibition period. For this reason, the start year of exhibitions is the crucial time-based property for this application.

4.1.3 Classification According to Concepts of Network Science and Graph Theory

It can be easily seen that the network is *undirected*, as links based on common exhibitions cannot have meaningful directions. Furthermore, the network is *weighted*, as we can store

the number of shared exhibitions in the links as properties. Our network is *multivariate*, because both links and nodes contain properties. The network is also *dynamic*, as it changes over the course of the exhibition years. Finally, it is *geospatial*, because of the geographical relation to the exhibition locations.

4.2 Users

The end users of this application are art historians. To meet the needs of art historians, it is crucial to consider our findings in Chapter 3.2.1.

One of the most important findings, especially with regard to the amount of data to be visualized, is that art historians prefer to study smaller networks [SRC17, LABV18]. From a technical point of view, this is beneficial, because smaller networks can be easily displayed using SVGs without having to consider performance issues. However, from a conceptual point of view, it is challenging to display exactly the nodes and links which are needed in a visualization. Which nodes and links are “not important enough”, so that they can be excluded from the visualization? In the networks Saint-Raymond and Courtin [SRC17] studied, they suggested to remove nodes with no connections, for example. Burckhardt [Bur17] only considered links based on two or more shared conferences in his network projection.

From our point of view, however, such approaches are problematic. We believe that the study of lesser known and seemingly less relevant artists can provide new domain-relevant insights. For this reason, we consider it important to include a tool in the application that allows art historians to create arbitrary small-scale networks themselves based on their research interests. If required, these networks can be edited at a later point in time. If the created network happens to be very dense, there must be a possibility to reduce the complexity with the help of meaningful filters.

Another point to consider is that art historians may be less likely to work with technical tools than, for example, scientists in STEM disciplines. We therefore believe that a web application is a suitable architectural style for the prototype, as the application can be used conveniently with conventional web browsers. Furthermore, we want to avoid complex visualization layouts as well as interaction techniques, which may be less intuitive to use and require a very long training period.

With regard to accessibility, color blindness in particular must be taken into account. Since colors are often used in visualizations, problems may arise during operation. It is important to avoid a complete dependency on colors when displaying important information.

Finally, we consider it essential to adapt an evaluation procedure that includes art historians as participants. For this reason, we rely on the insight-based methodology presented by Saraiya et al. [SND05], which we will describe in detail in Chapter 7.

4.3 Tasks

Miksch and Aigner [MA14] discuss that it is often difficult to capture the tasks a user wants to carry out with the help of visual analytics solutions. For our application, we mainly focus on the tasks that are defined by the DoME project team. According to the project team, the data set should, among others, help to investigate the “geography and networks of modern painting” [DoMa]. Specifically, it should help to answer the questions “Where (in which cities, in which institutions), when, and with whom did artists exhibit?” [DoMa]. In our application, we will visualize the network of artists and focus on shared exhibitions, while individual exhibitions of artists are *not* considered. Finding single exhibitions of a particular artist can already be done easily on the DoME website [DoMb].

Despite this already existing task specification, it is important to capture the tasks more precisely with the help of a task taxonomy. Miksch and Aigner [MA14] discussed different taxonomies, especially with respect to time-oriented data. For our application, we rely on a task taxonomy for graph visualizations proposed by Lee et al. [LPP⁺06]. Based on the low-level visual analytics tasks described by Amar et al. [AES05], they developed a taxonomy consisting of four task categories, which should support designers and evaluators in their work. The categories include topology-based tasks, attribute-based tasks, browsing tasks, and overview tasks.

Discussing tasks with respect to these categories also allows us to contrast node-link diagrams with matrix-based visualizations. In our SLR, we identified matrix-based visualizations as a commonly used and useful [GFC05, ABHR⁺13] alternative to node-link diagrams, which is why we want to incorporate this visualization layout in our application as well. In particular, with the help of the task taxonomy, it can be concluded that matrix-based visualizations and node-link diagrams complement each other well for the fundamental tasks we identified. In Chapter 7, we compare these two layouts with each other empirically in an insight-based evaluation approach.

Topology-based Tasks

We found in our SLR that topology-based tasks are very valuable in the art history domain. Among them is the study of various centrality metrics [SRC17, LV20, Ric17, Bur17, McC19] and the description of structural features such as communities [Ric17, McC19, NVCD22, Bur17] of social networks. In our application, studying topological features is also important in order to identify important exhibiting cities or key artists in artists’ networks. For simple topology-based tasks, matrix-based visualizations proved to perform particularly well [GFC05].

Attribute-based Tasks

Attribute-based tasks are of special importance in our multivariate network, because we can encode geographic information in the links using network projection. The insight that two artists exhibited with each other becomes particularly valuable if the cities of the

shared exhibitions can also be named. Based on former task-based evaluation approaches [GFC05, ABHR⁺13], we believe that matrix-based visualizations can be beneficial for these types of tasks, as they highlight the links more prominently than node-link diagrams. Attributes are also stored in our nodes. For example, it can be examined how strongly female artists are involved in a network and with which other artists they exhibited.

Browsing Tasks

Browsing tasks, which include the tracing of paths, are in our view negligible for the exhibition network we want to visualize. Links are based on shared exhibitions, which is why the focus is primarily on adjacent neighbors, rather than distant nodes. As we have already noted, node-link diagrams are generally superior to matrix-based visualizations in path-finding tasks [GFC05].

Overview Tasks

Overview tasks, which according to Lee et al. [LPP⁺06] can also include simple topology-based tasks, are important in our application. For example, network visualizations can easily show whether artists exhibited together at all. Such overview tasks, while simple in terms of network visualizations, are much harder to complete with data in tabular form and therefore provide an example for the utility of network visualizations in general. We think that node-link diagrams using a force-directed layout are particularly valuable for overview tasks, as the topology is well highlighted by the arrangement of nodes on the surface.

Other Tasks

Lee et al. [LPP⁺06] also explicitly provide examples of high-level tasks that are not covered by these categories. These include the question of whether the network has changed over time, which is clearly an important task in our application. For comparison tasks, matrix-based visualizations can be of great use [ABHR⁺13]. From our point of view, the symmetry property of adjacency matrices for undirected networks, specifically, can be utilized to compare networks of two years in a single visualization. This is because only one half of the matrix needs to be used to visually encode the links of a network.

Another example the authors discuss are tasks which require interpretation by the users. In our application, these mostly include tasks that can only be performed with the help of domain knowledge. For example, the high betweenness centrality of an artist can be explained in terms of his or her influence on artists' associations. We think that the existence of such domain-specific tasks underlines the importance of a more domain-centered view for the evaluation, as provided by the insight-based methodology by Saraiya et al. [SND05].

4.4 Summary and Resulting Visualization Layouts

4.4.1 Visualizing the Social Network

The artists' networks are created with the help of network projection, with which the links include information about the number of shared exhibitions, exhibition years, exhibition cities, and exhibition countries. We decided to include both the exhibition cities and countries to allow for different levels of detailed study of the locations. The years are needed to query for links of specific years, i.e., to display the dynamic network as a sequence of yearly networks. Finally, the number of shared exhibitions can be understood as link weights, which denote the quality of a relationship between two artists.

Thus, the resulting network consists of artists, represented as nodes, and the shared exhibitions, represented as links. To visualize this network in its basic form, we rely on two layouts as mentioned in the previous Section, namely node-link diagrams and adjacency matrices. To allow art historians to study the network in detail with respect to time, we visualize both the network in the entire period as well as the dynamic network, which consists of yearly networks from 1904 to 1916. By incorporating meaningful filters, we allow art historians to reduce the complexity of particularly dense networks. Coloring functions can be used to highlight node properties such as sex or nationality, but special attention must be paid to accessibility, i.e., it must be able to analyze the attributes despite color blindness.

4.4.2 Map

As with the application *Palladio*, we would like to complement the network visualization with a map, in our case to display the exhibitions. All data related to the locations and exhibitions could be encoded within the links of our geospatial network, including the exhibition titles as well as the coordinates. However, we want to avoid displaying too much data within a single visualization component. Thus, only some information about the geographic locations, namely cities and countries, is displayed within the network visualizations, while more detailed information about the exhibitions is displayed in a separate map component.

Implementation

5.1 Labeled Property Graph Model

Similarly to previous publications we found [SRC17, LV20, Ric17, Bur17], our social network is created using network projection, in our case, on the basis of shared exhibitions. Special attention must be paid to the performance of the application, since large exhibitions display works by hundreds of artists, and the DoME data set contains over 1,300 exhibitions. Thus, a projection based on shared exhibitions inevitably leads to an enormously large number of many-to-many relationships, i.e., many artists can exhibit with many other artists. Traditional relational database management systems (RDBMS) perform poorly when dealing with a large number of these relationships, which is why graph databases provide a more suitable alternative [HTL17].

The choice of a particular graph database depends on the underlying *data model*. Silberschatz et al. [SKS96] define a data model as “a collection of conceptual tools for describing the real-world entities to be modeled in the database and the relationships among these entities”. Two of the most popular *graph data models* are the *Resource Description Framework* (RDF) as well as the *Labeled Property Graph* (LPG) [PVC21]. While RDF is used predominantly in the *Semantic Web* for linked data [VGSVZ⁺19], LPG has established itself as an appropriate data model for graph analytics tasks in particular [PVC21]. In this data model, nodes and links can store attributes as key value pairs. A pioneer and well-known implementation of the LPG is *Neo4j*, which aims in particular for efficient storage, fast querying as well as quick traversals through the graph [Bar].

We think that Neo4j is in many respects the ideal database for a visual analytics application in the area of network visualization. Besides the performance advantages mentioned above, the intuitiveness in particular should not be underestimated. The multivariate network to be visualized is already stored in a corresponding data model. Thus,

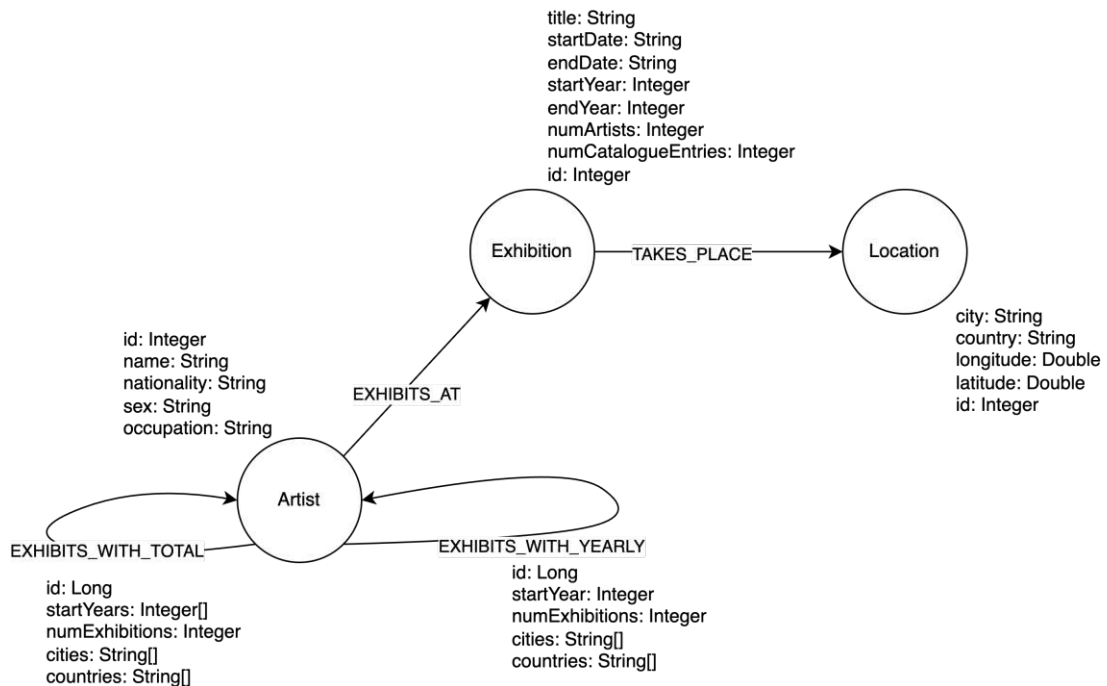


Figure 5.1: Final labeled property graph database model used for *Exhibitions Explorer*.

complex queries, which are sometimes necessary to prepare the data for the visualization, are kept as simple as possible. The concrete data model for our application can be seen in Figure 5.1. As it can be easily seen, the social network that is to be visualized is actually stored as such in the database. There are two implementation details which are worth noting. First, Neo4j only allows the storage of directed links. When querying with *Cypher*, Neo4j’s graph query language, link orientation can be omitted. However, care must be taken that the same link is not queried twice when the direction is not specified. The second implementation detail concerns the distinction between annual links and total, aggregated links. While it is also possible to reconstruct the aggregate connections based on the annual connections between artists, this would be far too time- and resource-intensive for the very large network we wish to store. Therefore, both connections are stored separately. Besides the social network of artists, locations and exhibitions are also entities in the database. These play an important role in the map component of the application. Furthermore, the social network was created on the basis of these entities using network projection. More implementation details will follow in the next Sections.

5.2 Application Architecture

Exhibitions Explorer is implemented as a web application. An overview of the architecture can be seen in Figure 5.2.

As mentioned earlier, we use Neo4j as our database, more specifically, *Neo4j Community Edition*, the open source variant of Neo4j¹.

A Spring Boot² application acts as a REST API. The connection to Neo4j is achieved via *Spring Data Neo4j*³, which implements an *OGM* (object-graph mapping) and provides convenient access to Neo4j. We rely on the *controller-service-repository pattern* for the implementation, which is particularly common in Spring Boot applications. This layered architecture provides a clear separation of concerns [Col21]. Each layer only accesses the layer below it. The repository layer represents the interface to our Neo4j database. The service layer is where the business logic is implemented. The controller layer serves as the interface to the frontend.

The frontend, which implements the user interface and thus also the visualizations, is an Angular application⁴.

5.3 Implementation Details and Approach

In this Section, the most important implementation details are presented in an overview. Please note that the implementation and technical details of *Exhibitions Explorer* are not the emphasis of this thesis, so only a high-level description of the implementation steps will be given. The source code of the backend and the frontend are available on GitHub⁵ under the GNU GPLv3⁶ license.

5.3.1 Database

The database of the DoME [DoMb] project was provided to us as a SQL dump. This was imported into a Neo4j graph database in a 2-step process. A comprehensive tutorial is provided by Neo4j [Neof]. In a first step, we identified the tables which provided us the properties for the nodes, which in our case were artists, exhibitions, and locations. The relationships between these nodes were already present in the SQL database as join tables, so identifying the links was very convenient. In a further step, the data was exported as CSV. With the help of Cypher, these CSV files could then be imported into the Neo4j graph database. The actual artist network was then created with the help of network projection, as already described. In a single link, exhibiting cities, exhibiting countries, years or year as well as the number of common exhibitions (i.e., the link weight) are stored. All this information can be taken from the exhibitions and location nodes as well as the corresponding *TAKES_PLACE* and *EXHIBITS_AT* links. Conceptually,

¹<https://github.com/neo4j/neo4j>. Accessed: 2022-11-27.

²<https://spring.io/projects/spring-boot/>. Accessed: 2022-11-27.

³<https://neo4j.com/developer/spring-data-neo4j/>. Accessed: 2022-11-27.

⁴<https://angular.io/>. Accessed: 2022-11-27.

⁵<https://github.com/boarder90/modern-exhibitions>. Accessed: 2022-11-27.

⁶<https://www.gnu.org/licenses/gpl-3.0.en.html>. Accessed: 2022-11-27.

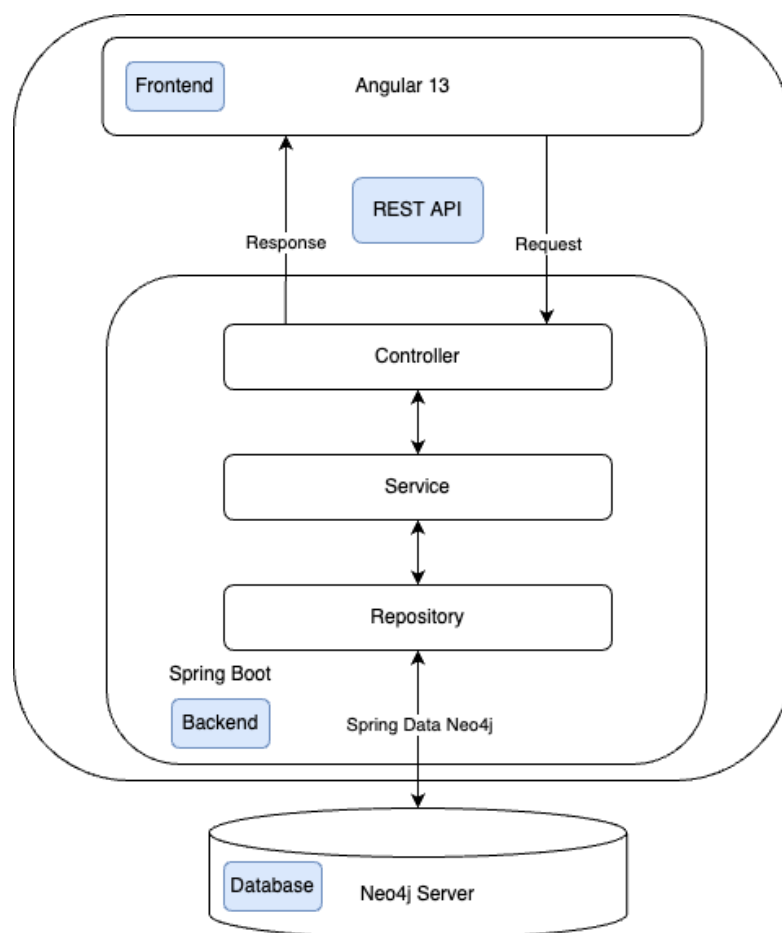


Figure 5.2: Web application architecture used for *Exhibitions Explorer*.

however, a link simply represents a joint exhibition activity of two artists.

A total of 14,858 nodes, including over 13,000 artists, and 7,672,691 links are stored in our database. The high number of links can be explained by the presence of large exhibitions, which lead to an enormous number of links in the course of the network projection. For example, if there is an exhibition with n artists, there are $\frac{n(n-1)}{2}$ links between these artists with a minimum link weight of 1. A higher link weight follows from the presence of multiple shared exhibitions.

The high number of links could be reduced by, for example, only storing links that encode at least two common exhibitions. However, as described in the previous Chapter, we wanted to avoid this at all costs, since from our point of view the often overlooked connections can provide valuable domain-relevant insights.

As we want to give users the ability to store the networks they create themselves, it is important to consider how the networks are stored. Neo4j currently does not support

the explicit storage of subgraphs [Neob], instead the corresponding links and nodes have to be marked with properties and labels. An alternative is to store them on the client side, which is also implemented by the Neo4j visualization tool Bloom in the free version [Neod]. We also decided to use the latter storage option. Thus, the artist IDs are stored as an array in the local storage of the user's browser. Based on these IDs, the network can then be queried from the backend.

Finally, it was also our goal to include algorithms that determine the centrality of nodes in our application.

Neo4j's Graph Data Science (GDS) Library⁷ can be used here. Unfortunately, we had to deal with technical limitations here as well. The library works with a special memory graph catalogue to efficiently implement the supported algorithms [Neoc]. Graphs have to be loaded into this catalogue using *graph projections*. Due to the dynamic nature of our applications, we need to be able to load corresponding networks in an undirected form (i.e., ignoring the orientations in the database) into this catalogue at any time using *Cypher projections*. However, for version 2.1 of the GDS library, loading undirected graphs into the catalogue is not supported using *Cypher* projections [Neoe]. For this reason, we could only use Neo4j's (weighted) degree centrality algorithm, which can be configured to ignore the orientation of the links. Other algorithms, including the one for determining the betweenness centralities of the nodes, require the graphs to be loaded as undirected graphs [Neoe, Neoa]. Since the source code for GDS library is publicly available under an open source license, an adaptation of the implementation could be done accordingly. However, this would not be within the scope of this thesis. We think that weighted degree centralities in particular can be very helpful to gain additional insights, as the link weights are taken into account when determining these and thus the quality of the links is also respected.

5.3.2 Backend

The main role of the backend is to fetch networks in a suitable structure from the database and transmit them to the frontend using GET requests. In a JSON structure, these are then sent to the Angular application. The networks are sent as JSON arrays of nodes and links with corresponding properties so that they can be easily processed in the frontend.

For the map view of the exhibition locations, we use the well-known *GeoJSON* standard⁸ to mark the exhibition locations. This increases the interoperability of our backend and allows it to be used with other applications.

⁷<https://github.com/neo4j/graph-data-science>. Accessed: 2022-11-27.

⁸<https://geojson.org/>. Accessed: 2022-11-27.

5.3.3 Frontend

We use various open source libraries in the frontend to simplify the implementation process. For the map, we rely on *Leaflet*⁹ using the map data from *OpenStreetMap*¹⁰ to display the exhibition locations on a map. For the network visualizations, we use *d3.js*¹¹, a well-known JavaScript library for creating dynamic data visualizations. For the user interface design we use *mdbootstrap*¹², a library that combines the solutions *Material Design* with *Bootstrap* to facilitate the implementation of responsive user interfaces.

The core feature of this application is of course the interactive visual exploration of networked data. Since we use well-known node-link and matrix-based visualizations, the implementation is straightforward with the help of *d3.js*. A large number of modules for network visualizations, which implement frequently used functionalities, can be found on GitHub. In particular, we make use of the *d3.js* implementation of the force-directed layout¹³. Furthermore, there are countless implementations of matrix-based visualizations as well as node-link diagrams on the web. We adapted two of them [Cra20, Bos20] in their basic form for our application, and modified and extended them to suit our requirements.

⁹<https://leafletjs.com/>. Accessed: 2022-11-27.

¹⁰<https://www.openstreetmap.org/copyright>. Accessed: 2022-11-27.

¹¹<https://d3js.org/>

¹²<https://mdbootstrap.com/>. Accessed: 2022-11-27.

¹³<https://github.com/d3/d3-force>. Accessed: 2022-11-27.

Visualization Design and Use Cases

The visualization application consists of three fundamental components:

- **Map component:** This component aims to display all exhibitions with some basic information about them on a map. Furthermore, users are redirected to the DoME[DoMb] website, if they want to obtain more in-depth details of a particular exhibition. The map component can also be used in combination with the dashboard component to display the exhibitions of an artists' network on the map.
- **Network creation component:** This component allows users to create networks with a maximum of 30 artists/nodes (the number can be easily increased in future), which can subsequently be examined in the dashboard component. The network can be created graphically using the ego networks of artists.
- **Dashboard component:** The dashboard component is used to study the created artists' networks. It has a fixed set of filters, which can be used to analyze the networks in more detail. Either a matrix-based or a node-link-based visualization layout can be used in the dashboard component. Both representations display the exact same information and support the same set of filters.

All visualizations support zooming. Furthermore, all node-link diagrams support drag & drop functionalities for the nodes in order to facilitate the analysis of the networks.

The components are described in detail in the following Sections.

6.1 Map

As in *Palladio*, we use a map in *Exhibitions Explorer*, in our case to display the locations of the exhibitions. The map component in our application is visually similar to the map used in the website of the DoME project¹, however, there are some notable differences, which are discussed in the following.

An overview of this component is shown in Figure 6.1. Figure 6.1a shows the navigation menu, which is present as a header in all components. Figure 6.1b shows the map. Larger collections of location markers are grouped together into one cluster. Here we make use of the *marker clustering* library² for Leaflet. Clicking on this cluster zooms to the according part of the map, displaying the individual locations or more fine-grained clusters. Single locations are displayed as blue circle markers, with the number of exhibitions displayed inside them. Furthermore, the size of the markers is proportional to the number of exhibitions. Less important exhibitions have the minimum size of 18 pixels.

As depicted in Figure 6.1c, buttons can be used to switch between a view of all years together and a yearly view. This functionality was also adopted in the dashboard component, which we will discuss later.

If one clicks on the individual location, the list of exhibitions for this particular location will be shown on the right, which can be seen accordingly in Figure 6.1d. Here, we focus on essential information: Only the exhibition title, the exhibition year, the number of catalogue entries, and the number of exhibiting artists are displayed. By clicking on *More*, one is redirected to the respective website of the DoME project, where comprehensive statistics and information on the respective exhibitions can be accessed.

The annual view can be seen in Figure 6.2. Clicking on the individual locations in this case only shows the exhibitions for the selected year. It is also easy to observe that London as an exhibition location appears larger on the map than Brighton, because more exhibitions took place in London this year.

It is important to mention that we only consider the start years of the exhibitions for the filters in our application. For example, if an exhibition takes place between 1911 and 1912, it will only be found on the map under 1911. We feel that this is necessary so that the years can be compared in a meaningful way. When studying the networks of exhibitions, we wanted to avoid that the same exhibition is shown twice as a link in the dynamic network view. Since the map view is intended to support the network view, we decided to be consistent here as well by including only the start year in the yearly filter. However, in the list of exhibitions displayed to the right of the map, the end year is also displayed accordingly if it differs from the start year.

¹<https://exhibitions.univie.ac.at/exhibition/map>. Accessed: 2022-11-27.

²<https://github.com/Leaflet/Leaflet.markercluster>. Accessed: 2022-11-27.

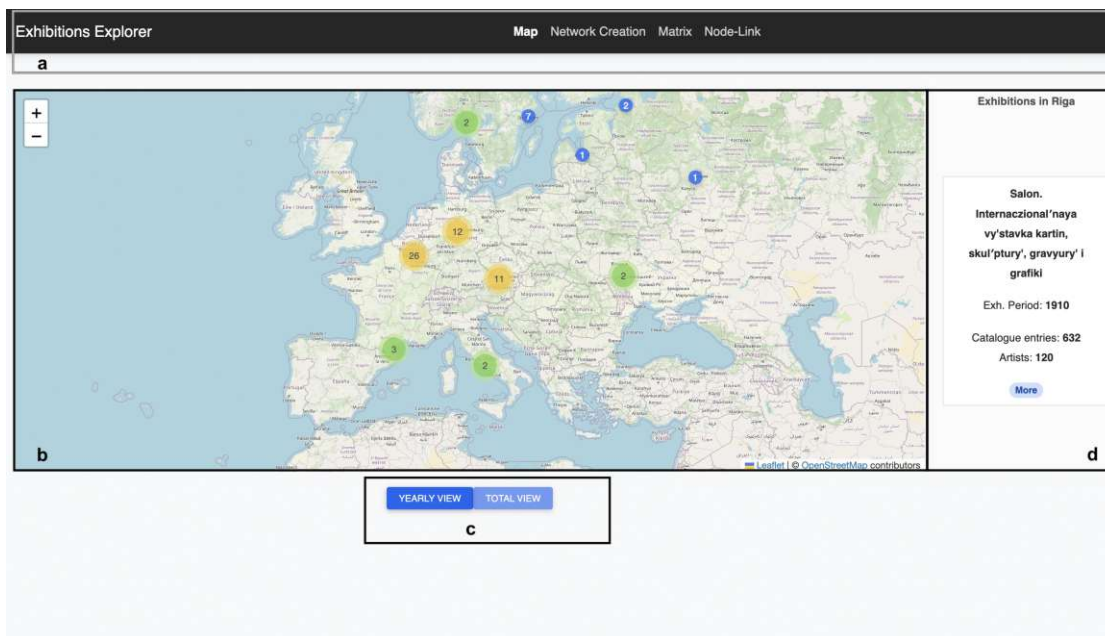


Figure 6.1: Total view in the map component, displaying all exhibitions from 1904 - 1916.

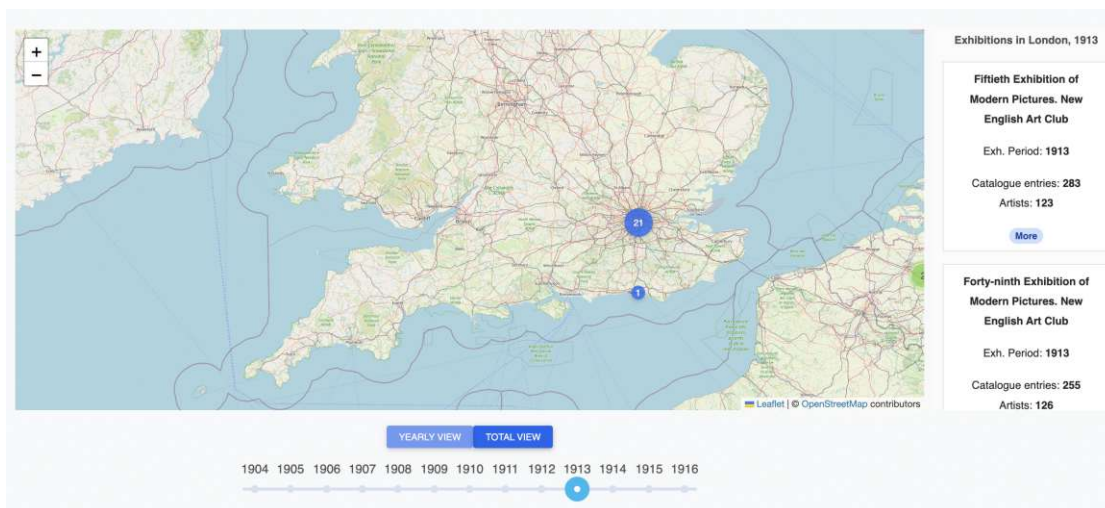


Figure 6.2: Yearly view in the map component. In this Figure, only exhibitions in 1913 are displayed.

6.2 Network Creation

One problem we addressed in Chapter 3.2.2 is the large number of nodes and links in many network visualizations, which make the analysis difficult. In fact, according to a study by Lamqaddam et al. [LABV18], smaller networks of less than 50 actors are already sufficient in an art historical context. Nevertheless, the question arises how the number of nodes can be reduced in a meaningful way without losing important information.

To this end, we want to allow users of the application to create networks themselves. This network creation component includes visualizations itself, so that researchers can identify interesting patterns and, based on this, create small networks of up to 30 artists in order to examine them more closely. The network creation component presented in this Section does not yet use any filters. We leave that to future work. Furthermore, the network creation component is not part of the evaluation, as the evaluation is dedicated to comparing the two visualization layouts used in the dashboard component.

6.2.1 Menu

As mentioned in Chapter 5, sub-networks are only stored on the client side. The created networks can be managed in a simple menu. Figure 6.3 shows the menu of the network creation component. Figure 6.3b shows how new networks can be created. First, users of the application must choose a unique name for the network. If a non-unique name is chosen, a corresponding error message will appear. Created networks can be managed below (see Figure 6.3c). They are displayed in a list and can be viewed, edited or deleted. This is displayed in Figure 6.4. If one chooses to edit the network, the user will be redirected to the selection view.

6.2.2 Selection

The actual selection of the network takes place in this view. The user can create networks in two ways: By simply entering the names of the artists or by exploring the ego networks of specific artists. An overview of this view can be seen in Figure 6.6. The user can search for artists by name (see Figure 6.6a). The user is assisted by an auto-complete function and can select the desired artist in a dropdown. Once the user has selected an artist, the ego network of this artist can be loaded. Figure 6.6e shows the loaded ego network of Peter Alma. The arrangement of the nodes happens automatically because of the force-directed algorithm mentioned in Chapter 5. The ego is highlighted in yellow. Coloring functions for all nodes are available on the left sidebar (see Figure 6.6b). The selected ego is also displayed on the left side of the view (see Figure 6.6c) along with the name of the artist, a short overview of his or her biography (occupation), the country of origin, and sex. The colors used for the countries of origin and sex are fixed. The colors displayed in the left sidebar are the same colors that are used for the coloring functions of the nodes. Should the user want to obtain more information about an artist, he or she can view it on the website of the DoME project [DoMb] by clicking on the *More* button. Finally, there is also an *Add* button, which is used to add the artist to

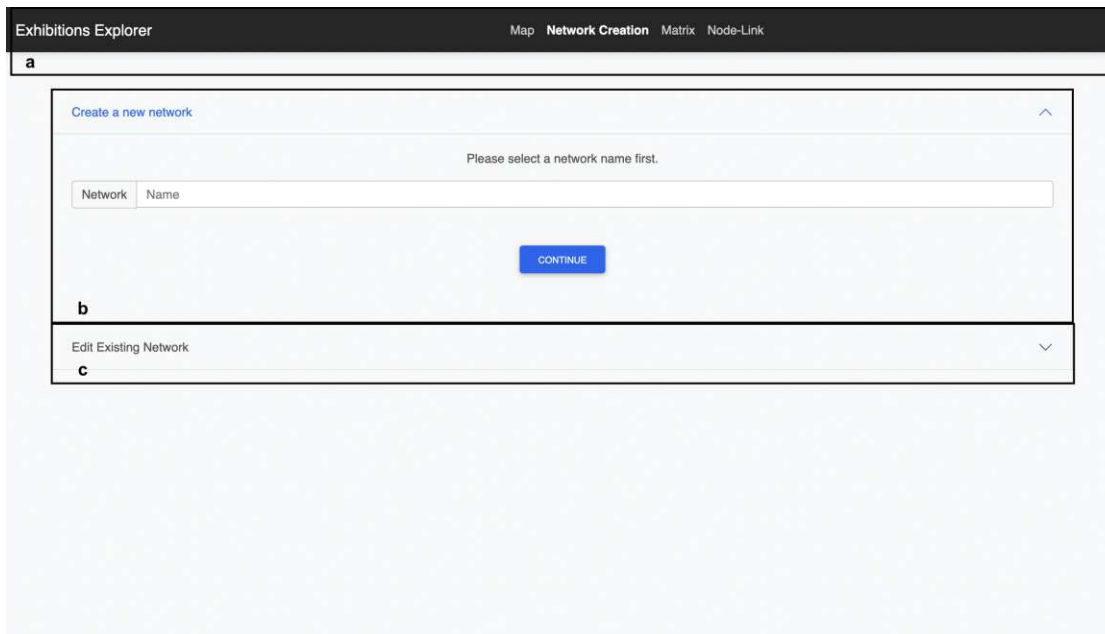


Figure 6.3: Overview of the network creation component.



Figure 6.4: Menu of the network creation component.

the created network. The number of currently selected artists is displayed at the bottom (see Figure 6.6d). Furthermore, the user can also go back to the menu view. Since all networks (to be precise, the artists' IDs) are only stored locally, no changes are lost. As no POST requests are used for the entire creation process, the selection is very fast and robust.

A key feature of the network creation component is the ability to navigate between ego networks. Once an interesting artist has been discovered, the corresponding ego network can be loaded by clicking on the *Ego Network* button in the sidebar on the right (see Figure 6.9f). This way, it is possible to navigate as far as desired to all accessible nodes of the initial ego network. To make the navigation more convenient, the *previous* ego studied is colored green. This can be seen accordingly in Figure 6.5. Here, the user navigated from Peter Alma to Jean Jovenau, so Peter Alma is now colored green while

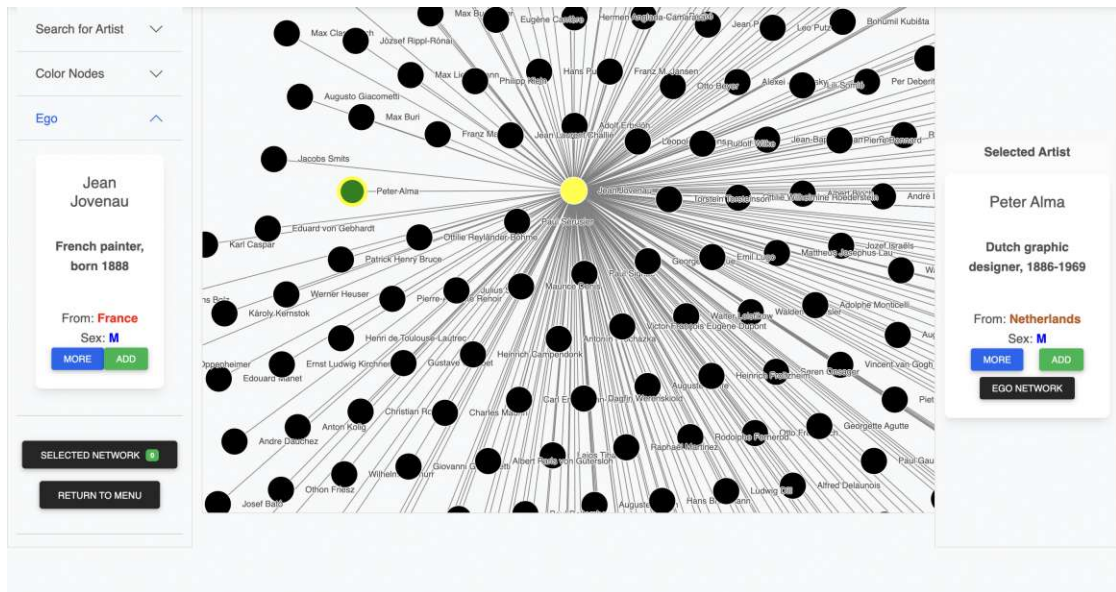


Figure 6.5: Navigation to another ego network. The previous ego is colored green, whereas the new ego is colored yellow.

the new ego Jean Jovenau is colored yellow.

The coloring functions can be seen in Figures 6.7 and 6.8. For sex-specific coloring, the colors blue and red are used to color male and female artists, respectively. Artists with no known sex are colored gray. Egos and previously visited egos remain yellow and green, respectively. When coloring by nationality, however, the special coloring of the ego and the previous ego is omitted, as the colors yellow and green are also used for nationalities. In this visualization design, multivariate visualization of the nodes can be observed for the first time. Different node properties in the LPG model are colored differently.

If the *Add* button is clicked, the artists will be added to the created network. Currently, a maximum of 30 artists can be added, however, this can be easily extended in future. If one clicks on the button *Selected Network* (see Figure 6.9a), the entire selection is displayed in an overview, as seen in Figure 6.9b. In this menu, artists can be removed from the selected network. Furthermore, the user can click on *More*, which again redirects to the according artist information page of the DoME project.

6.3 Dashboard

The main visualization component of this application is the dashboard component. In Figure 6.10a, it can be seen that the same component is reused for both visualization layouts, specifically, the adjacency matrix and node-link diagram. Therefore, with the exception of the slider (see Figure 6.10e), which is slightly different for the matrix-based visualization method, all the characteristics and features are identical for both

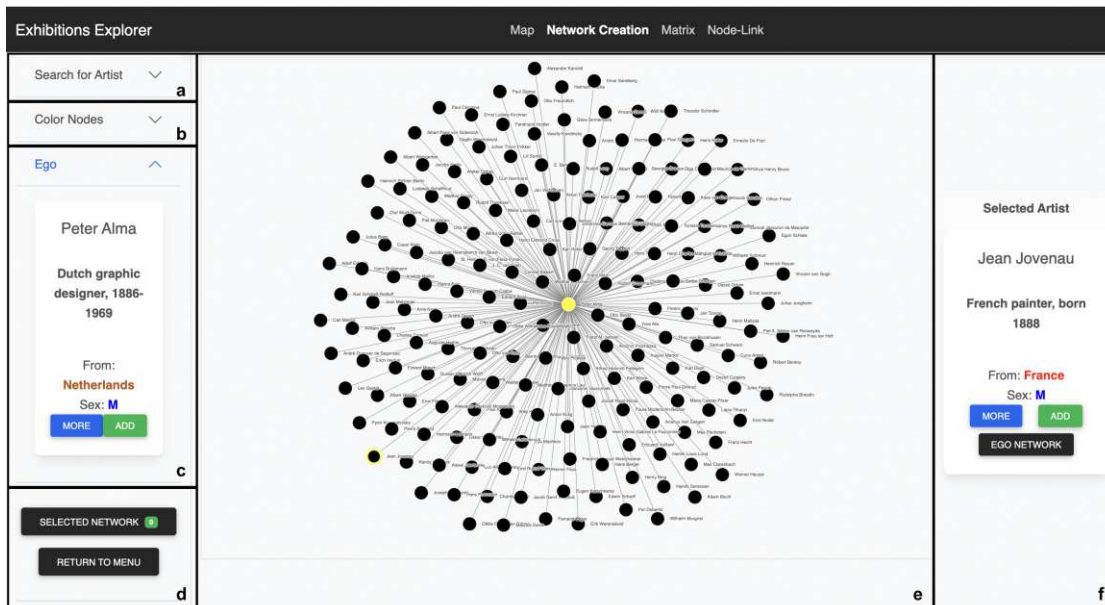


Figure 6.6: The network selection allows to create networks by exploring ego networks of selected artists.

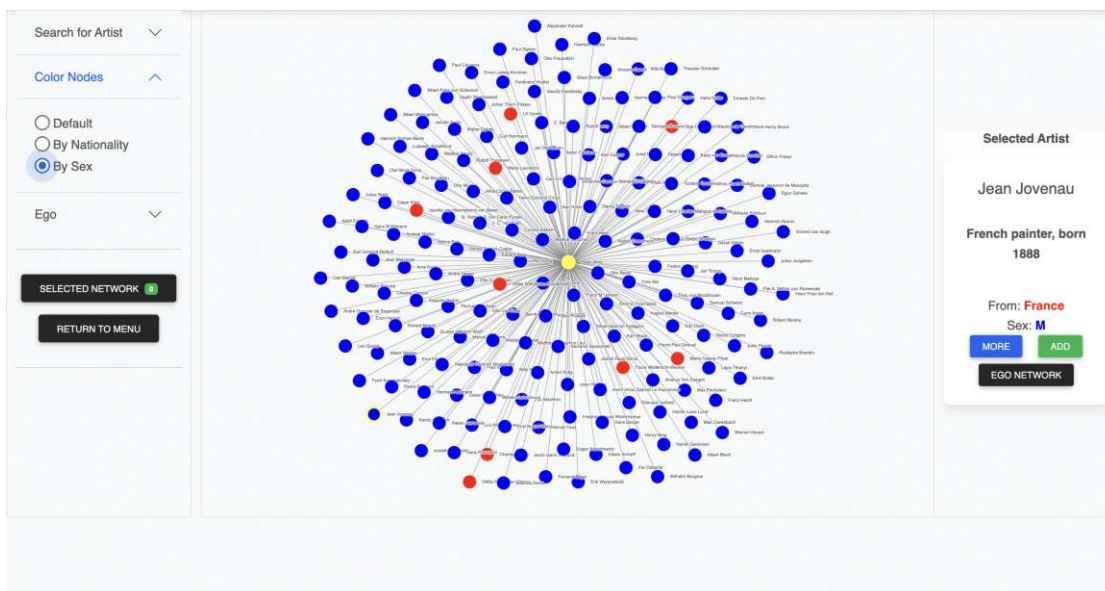


Figure 6.7: The nodes can be colored by sex.

6. VISUALIZATION DESIGN AND USE CASES

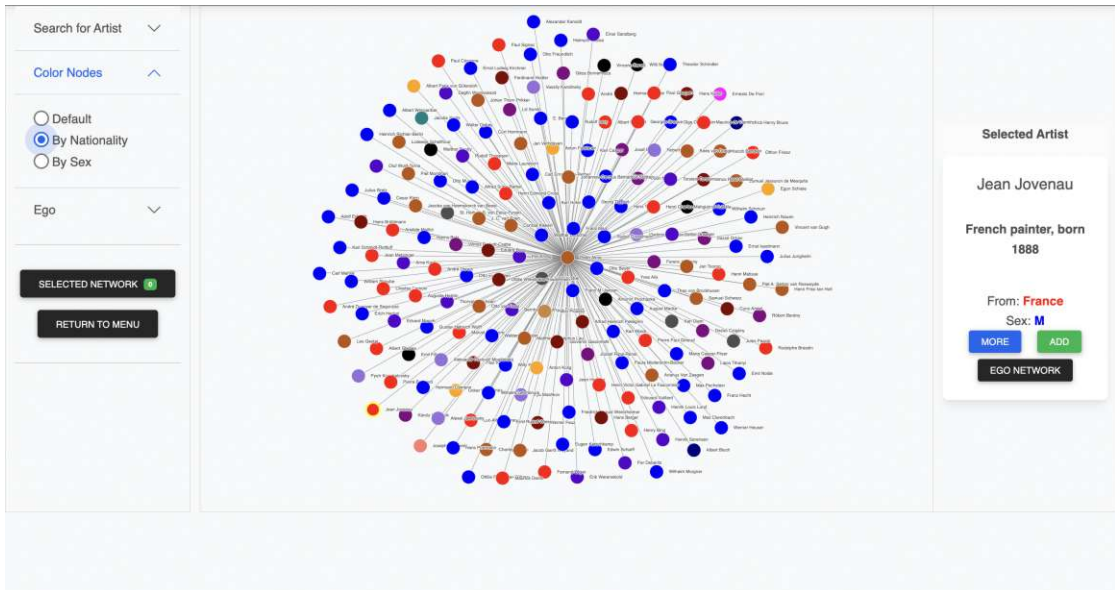


Figure 6.8: The nodes can be colored by (primary) nationality.

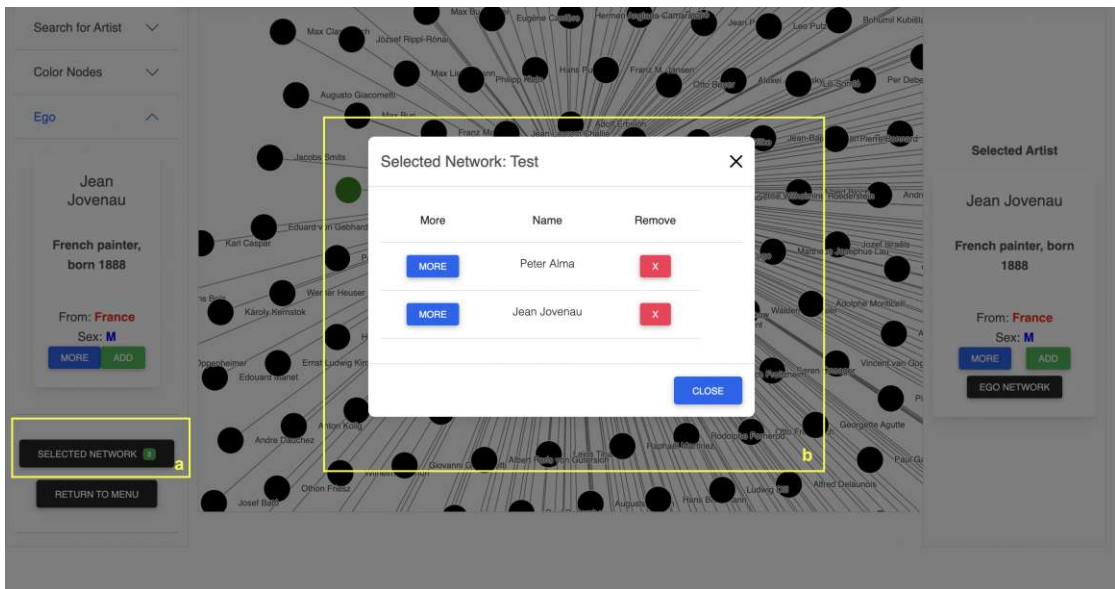


Figure 6.9: The modal displays all selected nodes of the created network.

representations. In Figure 6.10b, all interaction-based features and the selection menu are shown, which can be expanded as an accordion menu. The following options are available in the left sidebar:

1. **Network Selection:** Here, one can select a specific network by name that has been created in the course of the network creation process. The selected network is then loaded via a GET request from the backend and displayed either as a matrix or as a node-link diagram (see Figure 6.10c).
2. **Color Nodes:** This function is almost identical to the function in the network creation process, except that in the default view *all* nodes are colored black. As before, users can again color the nodes by sex or nationality. The node coloring can be combined with any other functionality, which leads to noteworthy use cases discussed in Section 6.6.
3. **Filter Links:** Users can filter links by exhibition country or exhibition city, so that a more compact network is displayed, showing only the common exhibitions in the selected locations as links. For the exhibition countries, the same colors are used as for the artists' nationalities. It is only possible to filter by a single country *or* by a single city, respectively.
4. **Centrality:** Here, nodes can be visually highlighted according to their degree centrality or weighted degree centrality. For the node-link diagram, this is done by increasing the size of nodes with particularly high centrality and decreasing the size of nodes with low centrality. In the matrix-based representation, the names of the artists are highlighted or de-emphasized by using font weight and opacity. This function cannot be combined with the city- or country-based link filter at the moment.
5. **Show on Map:** Here, the user can choose to (i) display the exhibitions encoded in a *specific link* on the map; (ii) display the *entire network* on the map. If the user clicks on a particular link between two artists in the dashboard component, all common exhibitions of these two artists can be viewed on the map using the component described in Section 6.1. Furthermore, the map component can also be used to display the exhibitions of *all links combined*. Thus, all exhibitions, where at least two artists of the network exhibited together, are displayed. Single exhibitions of artists are *not* shown, as they are not considered in the networks either. Regardless of the user's choice, the map component is always opened as a new tab. Thus, the user is not interrupted when exploring the network. In the map component, all previously presented functions in Section 6.1 are available again. For example, the user can filter the common exhibitions of two artists by year or view more information about a particular exhibition on the DoME project website [DoMb].

6. **Cities / Countries toggle switch:** This function is related to the right sidebar (see Figure 6.10d). If the user hovers the mouse cursor over a link, either the common exhibition countries *or* the common exhibition cities will be shown on the right. Thus, the switch can be used to regulate the granularity of the information displayed in the right sidebar. This is particularly useful if the user only wants to get a broad overview of the network’s geography.

The right sidebar (see Figure 6.10d) is used to dynamically display detailed information. Beside the common exhibition countries or cities, the exhibition year is also displayed when the network is viewed in the annual view and the user hovers over a link. Furthermore, the user can also hover over the artists (nodes) of the network. In this case, the name, sex, and nationality are displayed on the right.

In the sidebar, one can also see the colors used for a particular data point. This way, we avoid having to display a legend for the colors used in the component, which would take up a lot of space. Instead, the user can retrieve such information dynamically when needed.

Similarly to the map component, the user can access an annual view of the network using a slider. For the matrix-based view, the yearly view makes use of a special visualization method, which we will present in the next Section. For the node-link-diagram, the yearly view works like a filter, i.e., only links for a specific year are displayed. All previously introduced filters and functions can be used for *both* the total network and the yearly networks.

Images as well as examples of this component are shown in the following sections.

6.4 Matrix-based Visualization

The matrix-based visualization of an example network can be seen in Figure 6.11. This visualization is based on a simple adjacency matrix, where the different link weights are highlighted using saturation. Very color-intensive fields indicate that these two artists exhibited together particularly often. If the mouse cursor hovers over a cell, the field is highlighted in orange. The two associated artists, in this case, Kandinsky and Balla, are also highlighted in bold, while the names of other artists are “suppressed” as the opacity of the font decreases. It can also be seen that the exhibition countries appear in the right sidebar, along with the number of exhibitions in these countries.

One of the main contributions in this work is the matrix-based representation used for the annual view. Here, we rely on the favorable property that our network is undirected: Since adjacency matrices for undirected networks are symmetric, we can “split” the matrix and use one half each for two consecutive years. This division is visually emphasized by keeping the two halves spatially apart. The respective years are displayed at the edges of the two matrices. According to the classification of Beck et al. [BBDW17], our matrix-based visualization can be classified as a *hybrid* approach, because we make use of step-by-step animations using a slider, but always display two years side by side in a

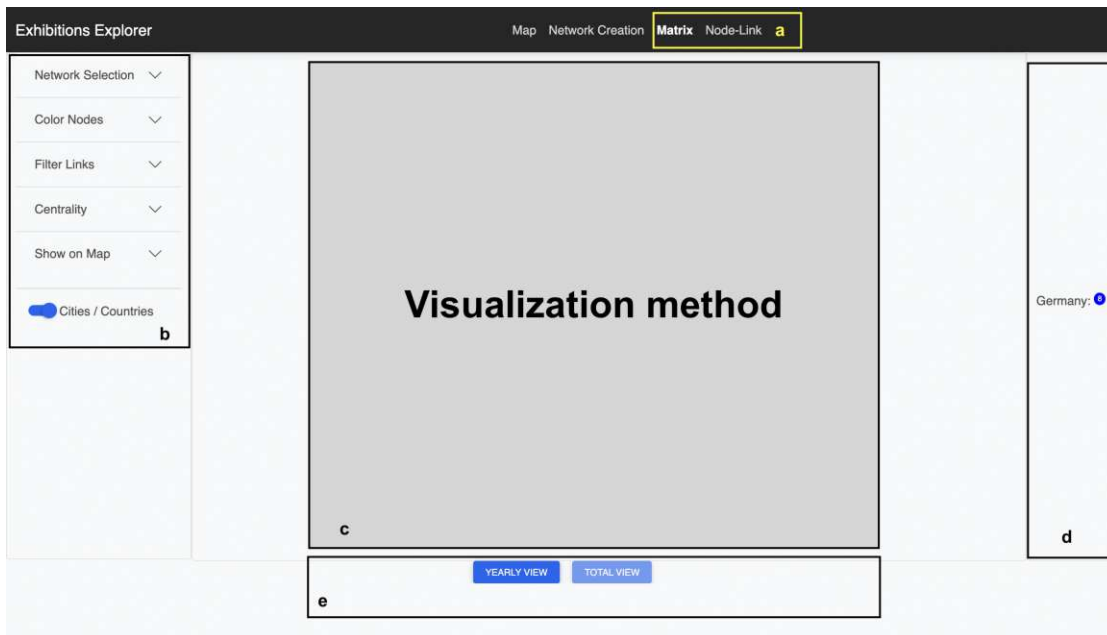


Figure 6.10: Overview of the dashboard component.

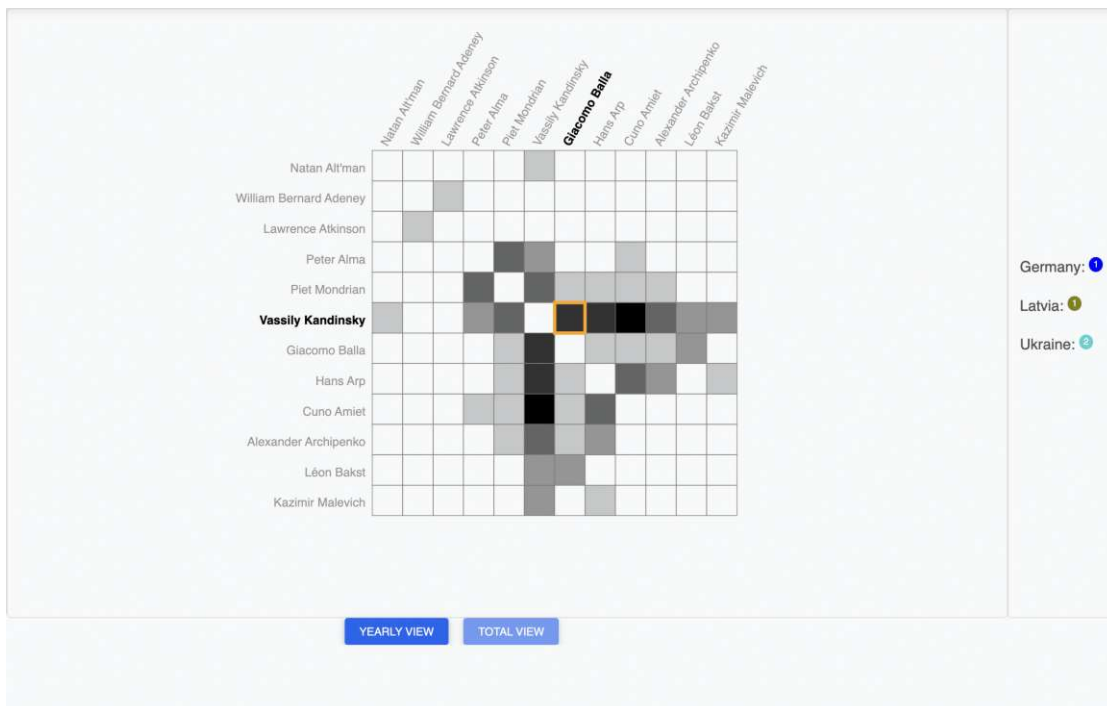


Figure 6.11: Matrix: Total network view, displaying the network in the entire period.

6. VISUALIZATION DESIGN AND USE CASES

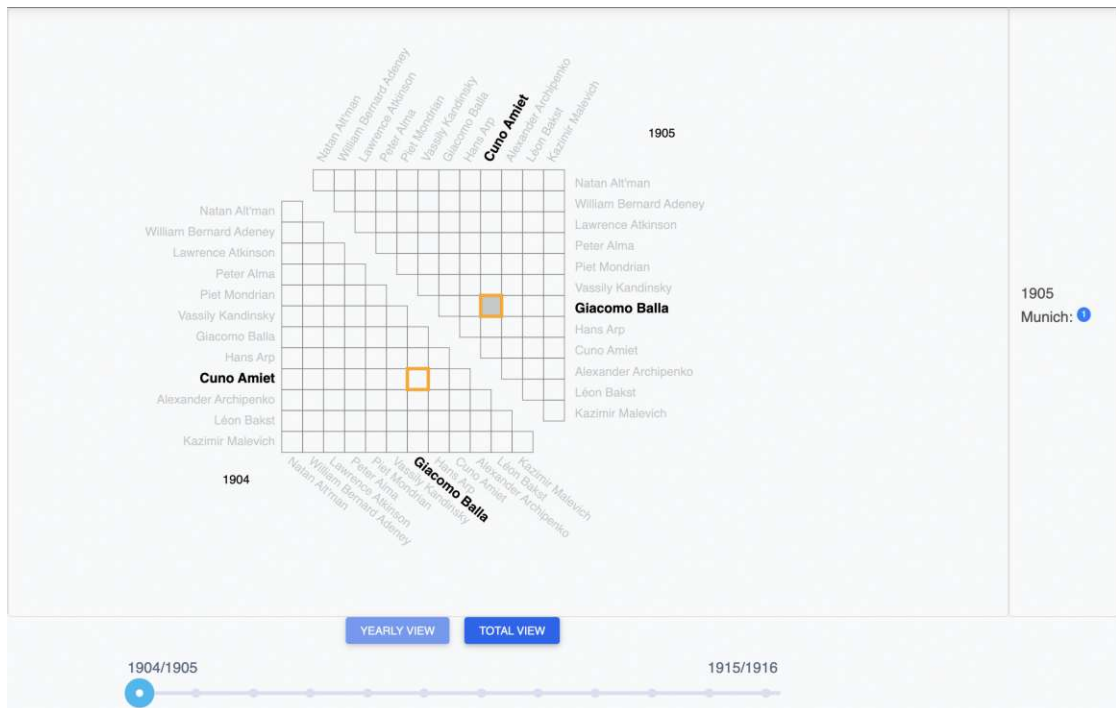


Figure 6.12: Matrix: Yearly network view, showing networks of two consecutive years.

small multiples approach.

An example of this visualization approach is shown in Figure 6.12. In this image, the mouse lies on a cell that shows the connection between Balla and Amiet in 1905. In this year, they had a common exhibition in Munich, which is shown accordingly in the right half of the matrix. In the visualization, the according names and cells of both years are highlighted simultaneously. We believe that this facilitates the comparison of both yearly networks. It can be seen that in 1904 all cells are white, so there were no common exhibitions in the entire network for this year.

The comparison of two annual networks is also facilitated by the fact that only one network is replaced step-by-step when using the slider. For example, when switching from the 1904/1905 view to the 1905/1906 view using the slider (see Figure 6.12), the network in 1904 is replaced by the network in 1906. The network in 1905, on the other hand, remains in the same place. This is illustrated in Figure 6.13. This makes it easier for users to comprehend how the network has changed over this three-year period.

We believe that this very simple and intuitive visualization approach significantly simplifies the comparison of annual networks to each other and can also be adapted for undirected networks in other domains.

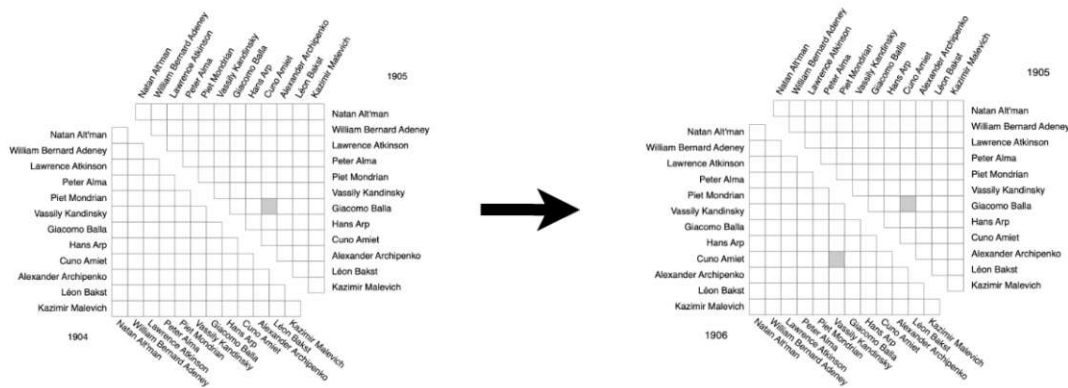


Figure 6.13: Change of the yearly network. It can be seen that the network of 1905 stays fixed in the visualization.

6.5 Node-Link Diagram

The node-link diagram relies on the popular force-directed layout to represent nodes spatially on the two-dimensional surface. This representation can be seen in Figure 6.14. To facilitate the analysis of the network, most nodes in the network are hidden when the mouse is placed on a single node, with exception of the direct neighbours of the node and the selected node itself. In Figure 6.14, the mouse is placed on Natan Alt'man, who only exhibited with one artist in the network, Vassily Kandinsky. All other nodes are hidden. This feature is especially important in case of extremely dense networks, which can easily arise due to our data model³. Furthermore, the nodes can be dragged and dropped freely inside the visualization area in order to mitigate denser areas of the network. Figure 6.14 also shows how information about an artist is displayed in the right sidebar. In the matrix-based visualization, this information would be displayed accordingly when the user hovers over an artist's name.

Link weights are indicated by the thickness of the lines connecting the nodes. Another option would have been to use link bundling techniques as discussed in Chapter 3.3.4, but we found our approach to be more clear. If a user places the mouse cursor over a link, all nodes except of the incident ones are hidden, as shown in Figure 6.15.

The yearly view of the node-link diagram is kept very simple. All annual networks are displayed individually for a single year, as shown in Figure 6.15 for an example network and the year 1910. Mental map preservation [CP96, AP13] is not explicitly taken into account here, since the force-directed layout in its basic version places a node on the surface based on the connections of that node. With respect to the classification of Beck et al. [BBDW17], this approach can be classified as a simple *animated* approach, because we make use of step-by-step animations using a slider.

³It should be stated that this is a common and popular feature for node-link diagrams, often found in open source-licensed implementations [Gó, Sua].

6. VISUALIZATION DESIGN AND USE CASES

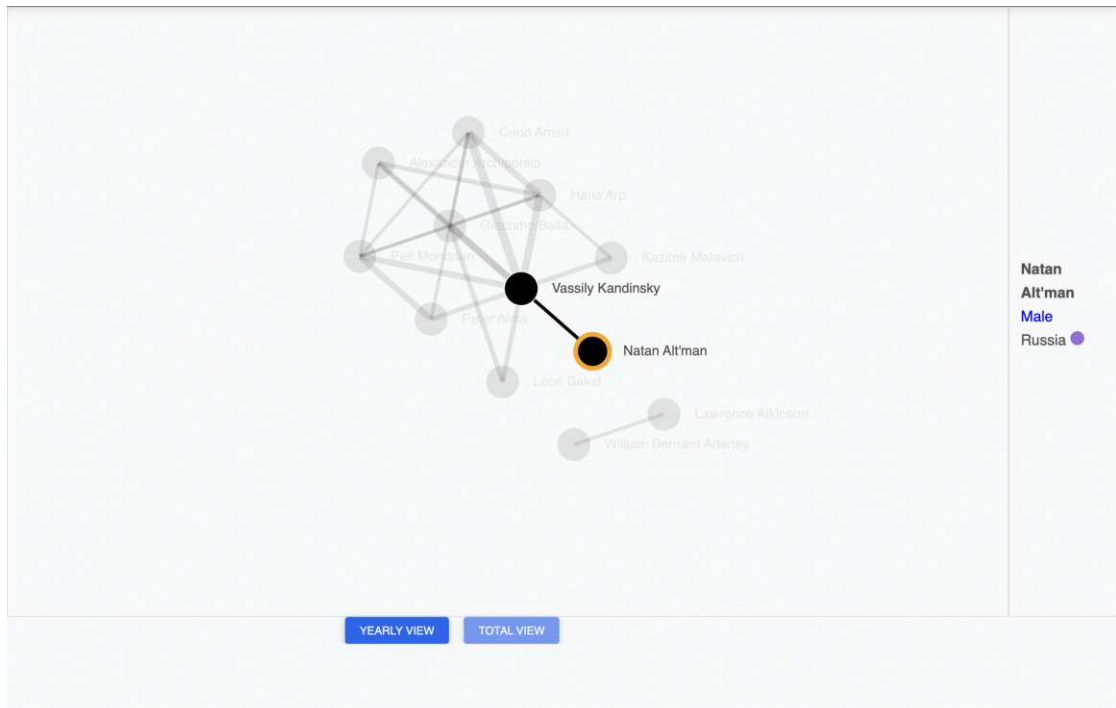


Figure 6.14: Node-Link: Total network view, displaying the network in the entire period.

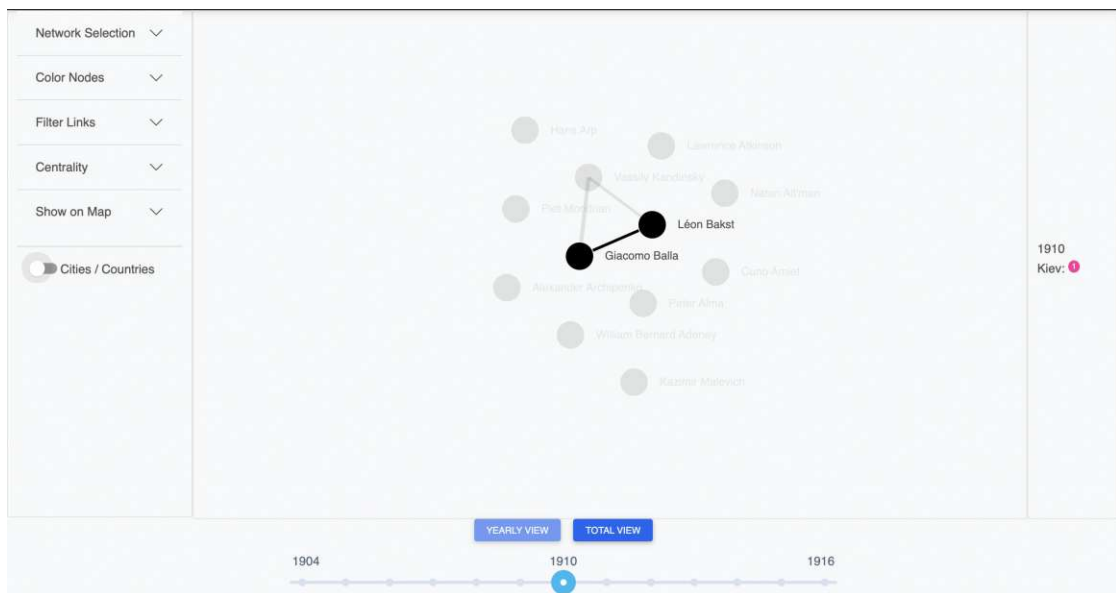


Figure 6.15: Node-Link: Yearly network view, displaying the network in 1910.

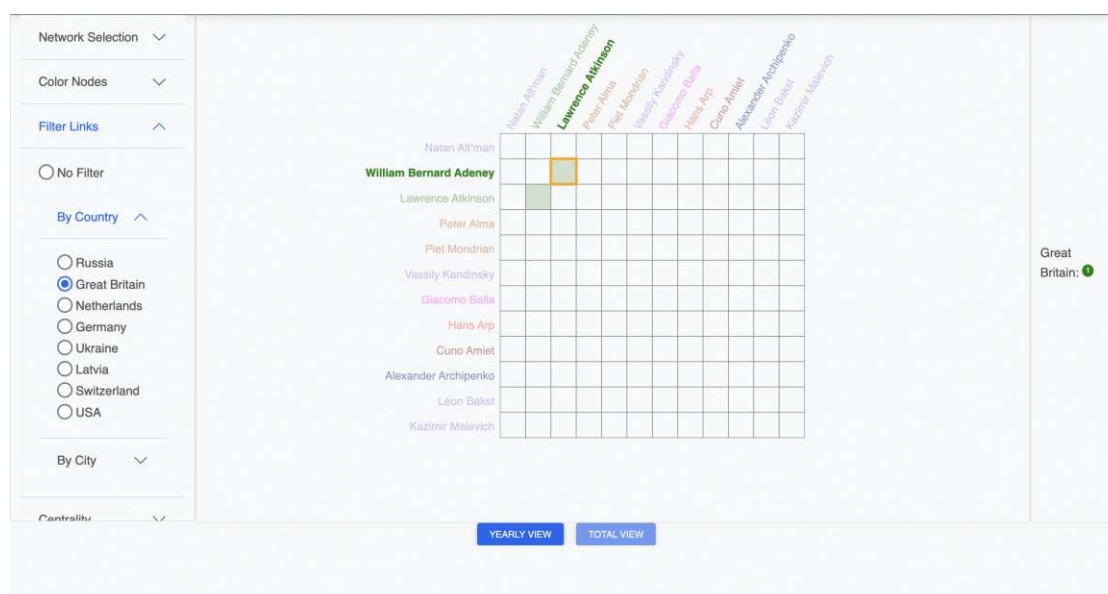


Figure 6.16: Combination of the filter and node coloring function. It can be seen that the British artists Adeney and Atkinson shared an exhibition in the UK.

6.6 Use Cases

In this Section, some selected use cases of the application will be presented, since the value of a visual analytics solution can best be demonstrated by domain-relevant examples. A network including some art historical background is also described in Chapter 7.2.1, which was subsequently used for the evaluation. Please note that the detailed art historical description of a network is not within the scope of this thesis.

Figure 6.16 shows how the node coloring can be combined with the link filtering. In this Figure, links were filtered by Great Britain, while the nodes were colored by nationality. As both functionalities use the same color, it can be observed that there is only a single link in this filtered network, which is based on a single British exhibition and this link also connects two British artists. Although two cells are colored in the visualization, this is due to the symmetry property of undirected networks.

The graph-theoretic concept of complete graphs has already been introduced in Chapter 2. In our application, such structures may indicate that all artists, which are part of a complete (sub-)network, share one common exhibition. To verify this, the user has to go to the map view and find the respective exhibition. Then, for example, the respective page of the DoME project website can be called in a further step, where all artists of an exhibition are listed.

If the network is filtered by an exhibition city and viewed in the yearly view, it can usually already be said with a very high certainty that a single exhibition connects all these artists.

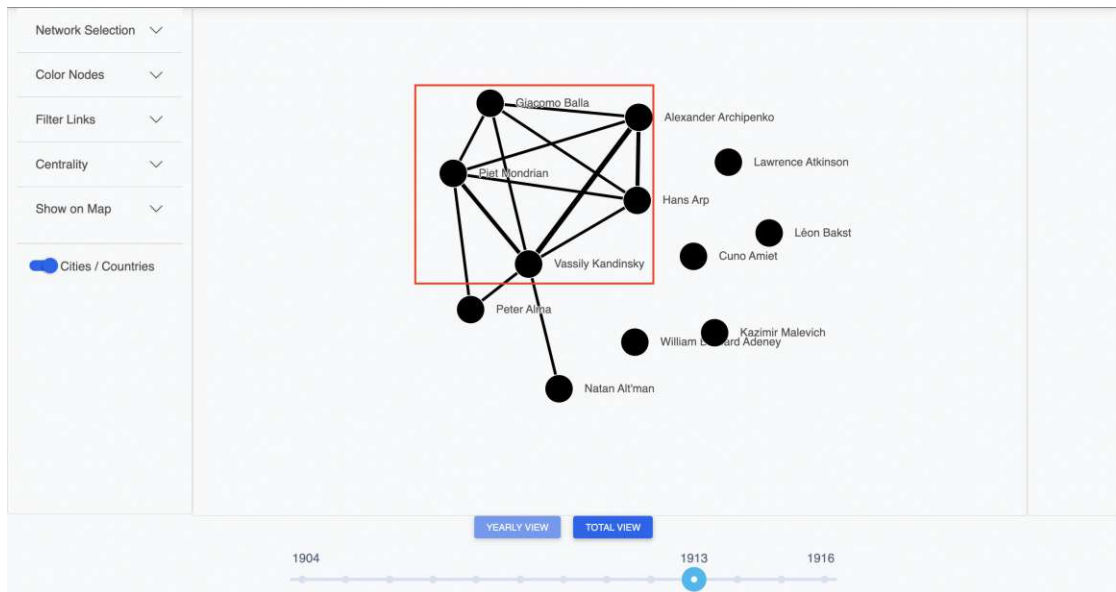


Figure 6.17: Complete (sub-)networks may indicate a key common exhibition.

Finding complete sub-networks is very simple using node-link diagrams, as it can be seen in Figure 6.17.

In Figure 6.18, it is shown how a particularly strong connection between two artists can be viewed in further detail on the map. If the user clicks on a link, which, in case of the matrix-based visualization, is marked by a cell, a green call sign appears in the left sidebar. The button, which opens the map in a new tab, contains the surnames of these two artists accordingly. The button for the entire network is always available if a network is currently being viewed.

Since centrality metrics play an important role in the publications we discussed in Chapter 3, a corresponding functionality is also present in this application. In case of the yearly view using the matrix-based representation, we benefit from the fact that two networks are displayed side-by-side, and thus the centralities of two years can be compared with each other in a single visualization. This can be seen in Figure 6.19. It can also be seen that the incident nodes of a connection are highlighted by a yellow background this time, because font opacity and font weight are used to display the centralities of the nodes.

6.7 Revisiting the Challenges identified in the SLR

In Chapter 3, many different visualization techniques were discussed for the representation of dynamic and geospatial networks. For our application, we focused less on complex visualization techniques and more on interaction as well as a precisely fitting data model to cope with the complexity of the massive social network.

6.7. Revisiting the Challenges identified in the SLR

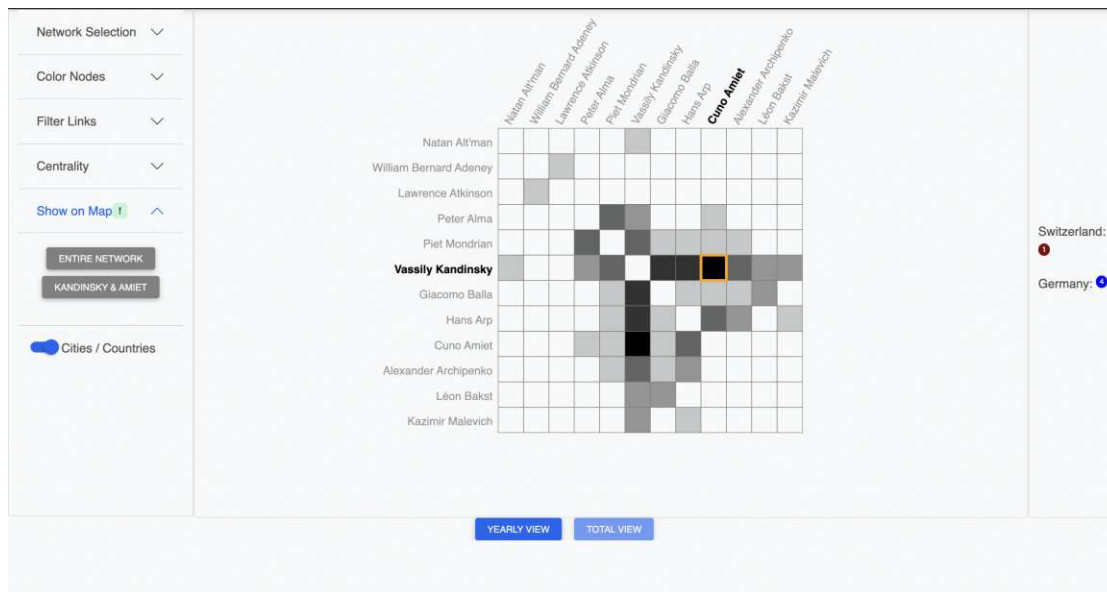


Figure 6.18: The map component can be (re-)used to display common exhibitions on the map.

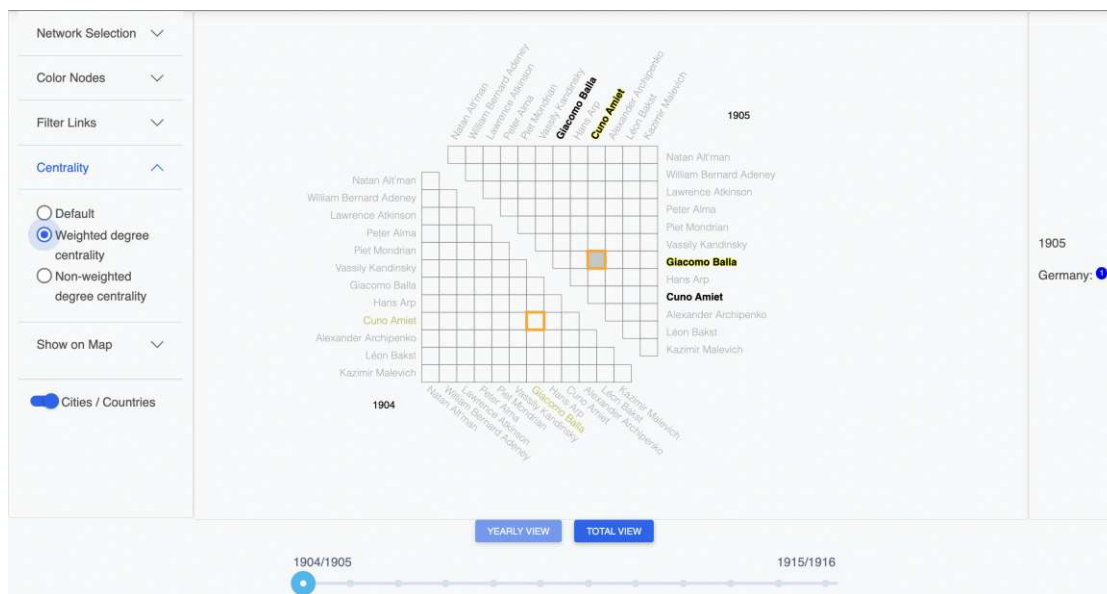


Figure 6.19: Weighted degree centralities of two networks in 1904 and 1905.

The user can switch from a dynamic network to a static, overall network at any time and thus control the granularity of the information displayed.

The interaction capabilities in visual analytics also make it easier to display the spatial nature of the network. Cities as well as countries encoded in the link properties are listed in a sidebar *if required*. The filter option also allows splitting the network by countries or cities. Kienle [Kie17a] pointed out that temporal characteristics of a network are often neglected in favor of geospatial ones in the art history domain. In our application, users are free to choose whether to focus on temporal aspects, geospatial aspects, or both, as the users can set the granularities as desired. The visualization of a complex spatiotemporal network is thus kept very simple with the help of basic interaction techniques.

In an interview Kienler [Kie17b] conducted with Druckner and Posner, it was emphasized that contextual information is often left out of network visualizations. In our application, for example, it is problematic that the importance of exhibitions, years, and places cannot be highlighted simply by using quantitative measures. For example, one exhibition year is not necessarily more important than another one, just because several exhibitions took place in that year. Rather, an art historical assessment must be made that examines the importance of particular artworks in an exhibition year. We think that this problem can partly be addressed through transparency: The user must be aware of the underlying data model and how data is displayed in the visualization according to the model. However, it should also be pointed out that certain issues in this respect cannot be solved on a technical basis. In visual analytics solutions, there is always the risk that a certain trade-off is made, which favors a clear representation over a more thorough one. Here, end users have to critically reflect on what they see based on the transparently made data model and explanation of the functionalities.

Evaluation

7.1 Introduction to the Insight-Based Methodology

Evaluating visual analytics applications is considered particularly challenging. In an article, Van Wijk [vW13] discusses the various issues that can arise when evaluating visual analytics solutions and gives an overview of existing approaches. According to the author, some of the major challenges include the large number and diversity of end users, the varying complexity of typical tasks performed on visual analytics applications and the enormous amount of heterogeneous data that are typically processed in such artifacts.

Van Wijk mentions lab experiments, which measure the accuracy as well as the time needed by users for completing predefined tasks, as a common evaluation method in the domain of Human-Computer Interaction (HCI). However, as visual analytics assists users in gaining new insights about a data set, different evaluation methods, which do not focus on isolated tasks, might be more appropriate. Unfortunately, measuring insights, as Van Wijk states, is considered a very difficult task, furthermore, the definition of insights is often rather vague.

To address some of these problems, the insight-based evaluation approach was introduced by Saraiya et al. [SND05] in an article. The publication is based on previous work [SND04] by the same authors, which describes a case study in the bioinformatics domain. Insights, as defined by Saraiya et al. [SND05], refer to “an individual observation about the data by the participant, a unit of discovery”. Each insight comes with a set of characteristics, such as, the domain value of the insight or the time it was made. These characteristics can later be quantified and encoded for analysis purposes. Chris North [Nor06], summarizes the key aspects of this evaluation approach as follows:

- An open-ended protocol is used, where users can explore the data set as desired, rather than relying on pre-defined tasks.

- The analysis of all insights happens in a qualitative manner. Users report all their insights, which are recorded during the evaluation. Then, metrics of these distinctive insights can be quantified and encoded for analysis purposes.
- To evaluate the relevance for the domain, domain experts should be used for the evaluation. They can also make claims which go beyond the pure observation of the data presented.

We will describe the entire evaluation approach as well as the adaptations of this approach in the existing literature in detail in the following Sections.

Despite the many advantages this evaluation methodology offers, it is important to point out that limitations exist here as well. Carpendale [Car08] compares different evaluation methods for information visualization based on three metrics as defined by McGrath [McG95]: *realism*, *generalizability*, and *precision*.

Realism refers to the proximity of the results to the actual context of use, generalizability refers to the usefulness of the results in other environments, and precision denotes the quality of the measurements. According to Van Wijk [vW13], realism is higher in methods, where experts share insights about the data set, especially when compared to traditional HCI-based methods. However, precision and generalizability may be an issue for the insight-based methodology. To increase the precision and comprehensibility of our results, we pay attention to a transparent and rigorous coding of the metrics, which was also highlighted as particularly important by North [Nor06].

7.1.1 Insight Characteristics and Categories for this Evaluation Protocol

The insight-based methodology of Saraiya et al. [SND05] needs to be adjusted for the evaluation of this application to ensure meaningful comparability of the results. A fundamental consideration is whether insights, i.e., individual observations of the data, or rather *insight categories* should be used for the analysis of the evaluation results.

Since this application displays networked data, it is very easy to accumulate a large number of insights by simply looking at each individual link of the network. Since each link displays unique information, i.e., the common exhibition of an artist pair, it is justifiable to count these observations as distinctive insights. If we only considered the quantity of single insights as a metric, the matrix-based visualization might have an unfair advantage: Because the links appear more prominent than in the node-link diagram, users might be tempted to look at the links and the encoded information in more detail. However, this does not mean that the matrix-based visualization is *better* than the node-link diagram, it is simply more *inviting* for certain types of insights.

Saraiya et al. [SND05] considered categories as insight characteristics, which allowed them to analyze the single insights in a more broad manner. In a following study, which

aimed to compare the insight-based evaluation approach to task-based evaluations, North et al. [NSD11] discussed the grouping of individual insights into distinctive categories for analysis purposes. The grouping helped them to analyze the performance of the participants per insight category and the number of insights per category.

We decided to adapt the approach by North et al. [NSD11] and analyze the insight categories based on the single insights made by the participants. For example, a participant might discover a link between two artists and report it accordingly. This insight would fall into a category “general claim about relationships of the network or the network itself”.

Furthermore, we will *score the categories* rather than the individual insights using the five-point Likert scale [PW10] to encode the domain value. This scale was also used in previous evaluation protocols [SND04, LBS⁺19, SNLD06, LNV⁺15]. In this approach, each insight falls into one category, while a category can contain multiple individual insights. Each insight category is awarded with a score from 1 to 3, depending on the domain value of this category. Using the previous example, the category “general claim about relationships of the network or the network itself” refers to very basic, trivial insights. Therefore, this category is assigned the value 1. The domain value was obtained based on comments of the participants during and after the evaluation, furthermore one participant assisted us in the coding process after all evaluations had been conducted. Analogously to Saraiya et al. [SND05], we award *hypotheses* with 4 or 5 points, depending on the complexity of the claim. In our case, we refer to hypotheses as statements that are made with the help of domain knowledge and go beyond the mere analysis of the data set. For simplicity, we will treat hypotheses as insight categories with high domain value in this protocol. Both correct and incorrect insights will be included in the categories and further analyzed.

In summary, the insight characteristics included in our evaluation protocol are *correctness*, *domain value* and *hypothesis*. Time was not considered as a characteristic, since trivial statements can be made instantaneously in the case of network visualizations. Furthermore, the time needed to make complex statements depends strongly on how the participant is exploring the data set.

We propose the following scheme to code insight categories, where each row should contain following entries:

- **ID:** Identifier for an insight category
- **Insight category:** Refers to the insight categories created on the basis of single insights. Hypotheses form insight categories themselves.
- **Domain value:** Refers to the domain value of an insight category or a hypothesis.
- **Counted correct insights per visualization method**
- **Counted incorrect insights per visualization method**

- Total counted correct insights
- Total counted incorrect insights

7.2 Evaluation Protocol

7.2.1 Data Set

Prior Considerations

A particularly challenging task was to create a suitable network for the evaluation. The network should have an adequate size and density considering the limited evaluation time window. Since the analysis of the network in our application goes beyond the study of topological structures, a rather compact network was deliberately taken for the evaluation. One problem was the choice of the thematic focus of the network. On the one hand, the evaluation deliberately relies on domain experts with prior knowledge, since the suitability of the visualization for creating or rejecting hypotheses is to be tested. On the other hand, a large imbalance in prior knowledge would lead to a distorted evaluation result. Accordingly, prior knowledge about expressionist painting (the main thematic focus area of the network), network analysis and network visualization was considered in the participant selection as well as in the analysis of the results.

In the evaluation network, both well-known and lesser-known artists were included. Besides the avant-garde movements Expressionism and Fauvism, which are also explicitly mentioned by the project team of DoME [DoMa], the artists of this network are also associated with other movements such as Impressionism and Symbolism. Particular attention was also paid to an interesting topological structure of the network.

Brief Historical Background

The artists of the network include Edvard Munch, Theresia Elisa van Hall, Anto Carte, Ernst Ludwig Kirchner, Anna Boch, Albert Servaes, Gustave de Smet, Pierre Paul Girieud, August Macke, Gabriele Münter, Otto Mueller and Isaac Grünewald. In Table 7.1, the associated movement as well as some basic historical information are presented in an overview.

Table 7.1: Brief historical background of the 12 artists displayed in the evaluation network.

Artist	Movement	Information
Edvard Munch	Symbolism, Expressionism [Aze15]	Influential Norwegian painter, who is also associated with German expressionism [Aze15].
Theresia Elisa van Hall	Unknown	Dutch painter, watercolorist and sculptor [DoMc]

Anto Carte	Symbolism, Expressionism, Naturalism [Car]	Antoine (Anto) Carte was a Belgian painter, who had a large number of exhibitions in Brussels [Car].
Ernst Ludwig Kirchner	Expressionism [AR10]	Influential German expressionist, who was a founding member of the artists' group "Die Brücke" [AR10].
Anna Boch	Impressionism, Pointillism [Bru05]	A popular Belgian painter who lived in Brussels [Bru05].
Albert Servaes	Expressionism [Eve]	A Belgian (Flemish) painter who was known for his religious art [Eve].
Gustave de Smet	Expressionism [DeS]	A Belgian (Flemish) painter, who is considered one of the main representatives of Belgian expressionism [DeS].
Pierre Paul Girieud	Expressionism, Fauvism [Gir]	A french painter, who is known for his strong ties with German expressionists [Gir].
August Macke	Expressionism [Ris17]	Macke was a German painter and one of the major representant of German expressionism. He is associated with the German artists' group "Blauer Reiter" [Ris17].
Gabriele Münter	Expressionism [Gun18]	A German expressionist painter, associated with the German artists' group "Blauer Reiter" [Gun18].
Otto Mueller	Expressionism [Sch66]	A German expressionist, who is associated with the artists' group 'Die Brücke' [Sch66].
Isaac Grünewald	Expressionism [Qva16]	A Swedish expressionist painter [Qva16].

Network

The network consists of 12 nodes and 23 links with differing weights. The network is displayed in Figure 7.1. It can be easily observed that artists of the Benelux countries share many exhibitions together. German artists share even more exhibitions together, as their links have very high weights. The strong connections of German artists can be explained by the existing artists' associations, as presented in Table 7.1. Belgian artists share a large number of exhibitions in the capital of Belgium, Brussels. German artists on the other hand share many exhibitions in Germany. Anna Boch and Pierre Paul Girieud both have a very high betweenness centrality and connect the two networks with

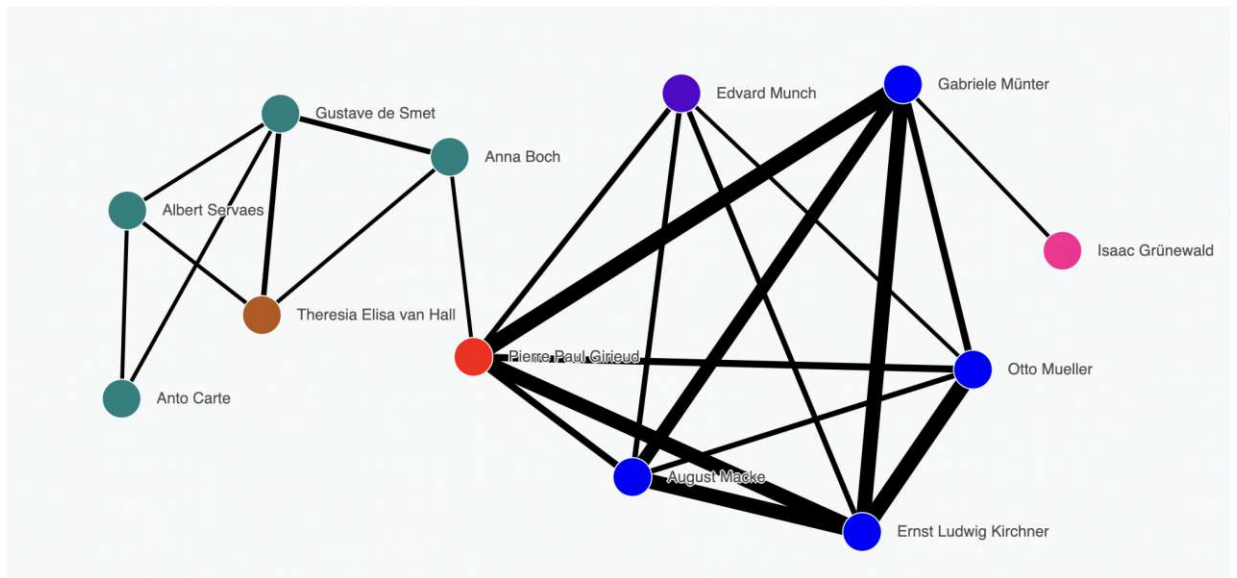


Figure 7.1: The network used for the insight-based evaluation, focusing on expressionist painters.

each other.

7.2.2 Participants

Four persons (three female and one male) were participating in the evaluation. All participants worked as mediators in the art domain and had an academic background in art history, holding a master's degree. The age mean was 49. No participant had experience in the field of network analysis or network visualization. No participant was familiar with the DoME [DoMb] data set. All participants reported to have some background knowledge in the area of expressionist painting.

7.2.3 Setting

All evaluations were conducted on an *Apple MacBook Air (M1, 2020)* laptop, which has a 13,3 inch screen with a 2560×1600 pixels resolution. The window for the application had a resolution of 1440×820 pixels. The participants were allowed to use the built-in touchpad of the computer or a mouse according to their preferences, however all participants decided to use a mouse. The keyboard was not needed at all for the evaluation. The evaluation session and the subsequent interview were recorded using a *Shure SM57* microphone. The evaluation language was German.

7.2.4 Introduction and training

The evaluation started with a short presentation on network analysis. Topics of this presentation included:

- What is a network?
- Link weights
- Node-link diagrams and matrix-based visualizations
- Undirected networks
- Centrality of nodes, with emphasis on degree centrality

The application was then presented. Due to time constraints, we could only present features of the application that were actually needed for the evaluation. This mainly included the dashboard component using one of the two respective visualization layouts and the map component. The entire training session took approximately 20-25 minutes.

7.2.5 Evaluation

In the subsequent evaluation, participants were asked to explore the network. Two participants each were assigned to a visualization layout (node-link diagram or matrix-based representation). The evaluation took a maximum of 30 minutes and was audio-recorded and transcribed for later analysis. In previous protocols, evaluation time was handled differently. Approaches include a lower bound [GGZL14], an upper bound [LBS⁺19], or no explicitly defined limits at all [SND04]. However, due to the schedules of our participants, the total session time had to be no longer than an hour, which is why decided to define an upper bound.

An important aspect of this protocol is that, similar to the evaluation protocol of Saraiya et al. [SND04], participants were allowed to ask questions about functionalities that they did not understand entirely.

The *think-aloud protocol* was used to collect the insights. This protocol has long been used in the psychology domain [Hop09] and it was also utilized as a data collection method in previous protocols of the insight-based methodology [LBS⁺19, SND04, LNV⁺15]. Gomez et al. [GGZL14] reported that they had experienced difficulties in counting insights that participants wrote down in textual form, which is another reason why we decided to use the think-aloud protocol. Participants were constantly encouraged throughout the session to report all their insights.

A decisive advantage of the think-aloud protocol is that we can also record spontaneous comments by the participants, which are not explicitly requested. Thus, we recorded occasional complaints or opinions on individual features, all of which we transcribed afterwards.

The evaluation was conducted in an open-ended manner. Before each evaluation session, Saraiya et al. [SND04] required the participants to ask initial questions that they hoped to answer with the help of the visualizations. Since our participants were not familiar with network analysis at all, we chose not to do so. Instead, we put together a guide with some general questions ourselves, which should facilitate the free exploration of the data set. This guide was shown once to the participants at the beginning of the evaluation, and furthermore, participants were not required to adhere to it. The guide is therefore more of an orientation aid and is intended to ensure that we can gather a wide variety of insights. The guide consisted of following open questions and is loosely based on the initial questions Gomez et al. [GGZL14] formulated for their protocol:

- Which countries / cities / years are to be highlighted? Which artists are to be highlighted?
- Which artist is particularly noteworthy? Which relationship is to be particularly emphasized? Can you make statements with temporal / spatial reference?
- Which exhibitions are particularly noteworthy? Can you find any on the map?
- Can you make general statements about the network? Are there any special structural features?

7.2.6 Semi-structured Interview

The evaluation was concluded with a semi-structured interview.

Semi-structured interviews are considered particularly common in the field of qualitative research. In these interviews, some questions or topics are determined in advance, with follow-up questions or spontaneous remarks being explicitly allowed or even desired. An important characteristic is also the flexibility in the phrasing of the questions as well as in the ordering [DV19].

Since the interview played a minor role in our evaluation protocol, we only prepared three topics for the interview:

- What is your opinion on the visualization method?
- Is there something you would like to change or improve in the existing application?
- Can you think of a particular visualization method (e.g., bar charts) and / or functionality that would complement the visualizations used in the application well?

The first two opinion-based questions were used to gather key impressions about the application as well as the visualization method. The last question was designed to collect

starting points for future work. Note that the exact wording of questions varied during the interviews, as we always followed up on the previous evaluation.

7.2.7 Transcription

The transcription and coding of the results is a very time-consuming work and one of the major issues of this evaluation method. We transcribed all recordings manually.

As noted by McCullin [McM21], there is no established “set of rules” on how to transcribe spoken text. A rough differentiation is that between *naturalized* and *denaturalized* transcriptions, as discussed by Bucholtz [Buc00]. The former “conforms to written discourse conventions” [Buc00], while the latter retains all details of oral communication. We opted for a naturalized transcription approach since linguistic details do not play a significant role in both the evaluation and the subsequent semi-structured interview. In particular, we did not transcribe strong stuttering; furthermore, we always translated dialects or strong colloquialisms into standard German.

We created three transcription files. The first file includes all the insights as reported by the participants. The transcription is based on the protocol described by Ledesma et al. [LBS⁺19] and includes following information:

- **Participant ID:** Refers to one of the four participants.
- **Visualization method:** Refers to one of the two visualization methods considered in this evaluation.
- **Insight:** Textual insights as recorded in the audio files. *Only* insights, whether correct or incorrect, were transcribed in this file.
- **Correctness:** Correctness was also captured in this transcription file to facilitate the analysis.
- **Time stamp:** While we do not consider time stamps to be an important metric for this particular protocol, time stamps were still captured to facilitate the later analysis of the insights.
- **Insight category:** To facilitate the later analysis of the insights, the insight category for each insight is also noted in this transcription file.

Spontaneous comments or complaints during the evaluation were transcribed in a separate file. In this file, we transcribed the comments as recorded and assigned them to participant identifiers and a time-stamp.

Finally, the semi-structured interview was transcribed as recorded, using the naturalized approach mentioned before. All statements made by the interviewer as well as

the participants were again assigned a time stamp and an identifier.

All transcriptions can be found in the appendix of this thesis.

7.3 Results

7.3.1 Insights

Insight Categories

We counted a total of 204 insights, with nine additional exact duplicates, which is why we removed them. Not removed were original insights, which were refined afterwards. For example, if the presence of a link was mentioned first, and then further specified with respect to the exhibition location, this was counted as two different insights. On the other hand, if a precise statement was made immediately, the insight was counted only once. Removing less precise, original insights would have made the analysis more subjective, difficult to understand and it would have increased the risk of errors; moreover, keeping the previous, less precise insights also allowed us to track whether or not participants were tempted to specify their insights in more detail when using a particular visualization layout.

We identified following 14 insight categories, which are shown in Table 7.2 (the reported hypotheses were translated and paraphrased). The categorization shows similarities to the graph taxonomy proposed by Lee et al. [LPP⁺06], but our categories are based on spontaneous, not explicitly requested insights, rather than on general, well-defined tasks. For a more detailed overview including the counted insights, please refer to the appendix.

Table 7.2: Insight categories, which are based on the 204 recorded insights.

ID	Insight Category	Domain Value
1	General claim about relationships of the network or the network itself	1
2	General claim about relationships of the network or the network itself with respect to the year and / or location	2
3	Complex, comparative claim about relationships of the network over the entire period	3
4	General claim about the artists of the network	1
5	Complex, comparative claim about the cities / countries of the network	3

6	Identify if two artists had one or more common exhibitions	1
7	Identify if two artists had one or more common exhibitions with respect to the year and / or location	2
8	Complex claim regarding the topological structure of the network	3
9	Claim about the degree centralities in the network	2
10	General claim about exhibitions or exhibiting cities (map)	1
11	Complex claim about the exhibitions over the entire period (map)	3
12	Hypothesis: While there are only a few female artists in the network, their importance grows over time as they become more densely connected. In order for them to get the chance to exhibit at all, the support of their male colleagues was probably important.	5
13	Hypothesis: Even in the time of origin, expressionists were already known outside the important hubs.	4
14	Hypothesis: The connections of the network are formed due to artists' groups and associations, such as "Brücke" and "Blauer Reiter".	4

Insight categories 1 to 9 refer to the network visualizations. Insight categories 10 to 11 refer to the map-based visualization, which was the same regardless of the network visualization layout used. Finally, insight categories 12 to 14 refer to the hypotheses made.

Incorrect Insights

We recorded a total of 11 incorrect insights, meaning that 94.6% of all recorded insights were correct.

From our point of view, this is a good result, since the participants, as already mentioned, had no previous knowledge in the fields of network analysis and network visualization. In addition, both our training time and our evaluation time were very short. It must be said, however, that there were some more incorrect insights, which were corrected shortly

Table 7.3: Incorrect Insights. Grouped by Participant ID, Visualization Method and Insight Category

Participant ID	Visualization Method	Insight Category
1	Node-Link (Map)	10
1	Node-Link (Map)	11
2	Node-Link	7
2	Node-Link	1
2	Node-Link	1
3	Matrix	6
3	Matrix	2
3	Matrix	2
3	Matrix	7
3	Matrix	2
3	Matrix (Map)	10

afterwards and independently by the participants. These were counted as correct insights accordingly. Furthermore, there were a few inaccurate statements, which we also counted as correct. An example of an inaccurate insight in this regard is that one participant immediately recognized and described Ludwig Kirchner's four high-weighted connections with Macke, Münter, Mueller and Girieud, but did not address the connection with Munch. The according insight was counted as correct, even though Munch was not mentioned until later in the evaluation. Since such insights are not *obviously* wrong, we decided to count them as correct in case of doubt - previous protocols also followed a similar approach [NSD11].

We were unable to find a pattern among the identified incorrect statements. Most of these statements were probably made too hastily. For example, one participant meant that there are *a total of* eight common exhibitions between two artists, and overlooked the one additional common exhibition in the Czech Republic.

Two incorrect statements were made by a participant who used the node-link diagram. Here, the presence of unconnected nodes was interpreted as artists exhibiting in a city, but *not together* with other artists. Accordingly, the participant was then highly surprised that there were exhibitions of German artists in Brussels. It is therefore important that we make it clearer in the future that this application solely examines artists' networks on the basis of common exhibitions and not individual exhibitions by artists.

An overview of the incorrect insights can be seen in Table 7.3. As already mentioned, we could not find any pattern worth mentioning.

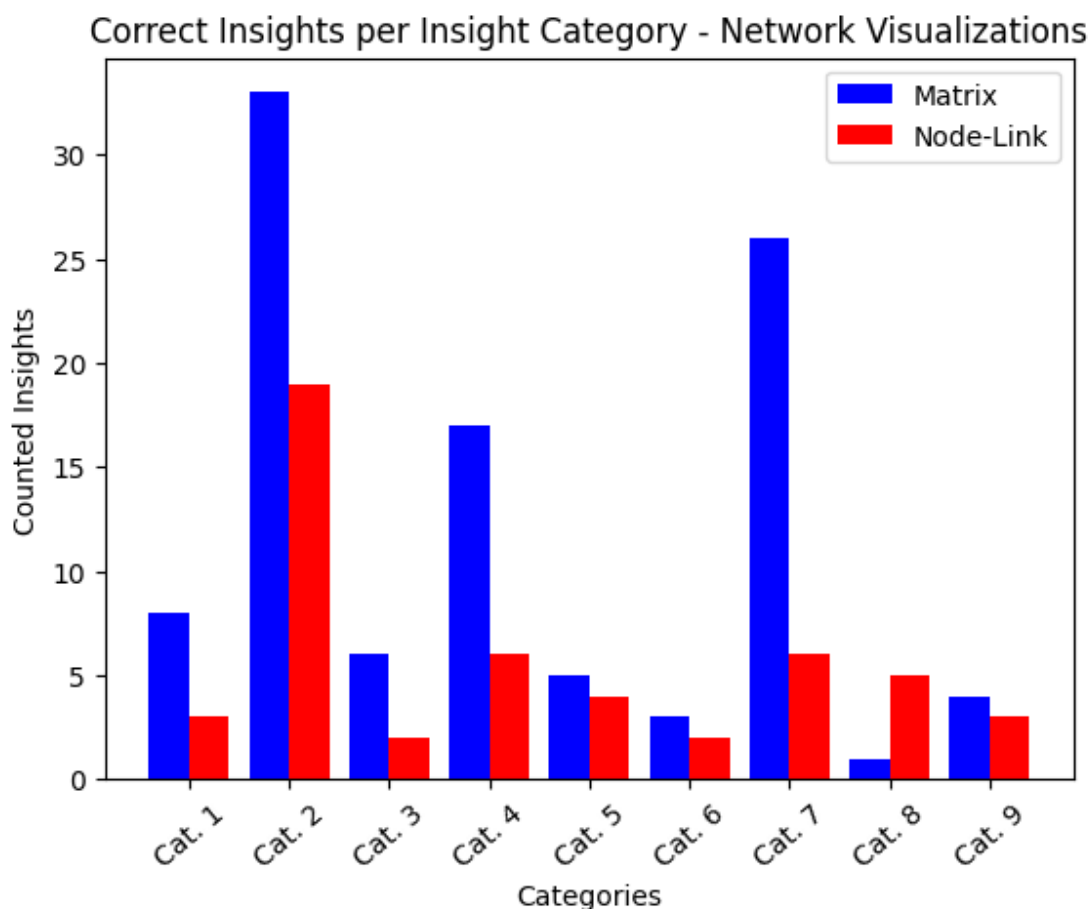


Figure 7.2: Correct insights gained using the network visualizations.

7.3.2 Correct Insights Using the Network Visualizations

Insight categories 1 to 9 refer to the network visualizations, which is why it makes sense to analyze them together. Figure 7.2 shows the collected correct insights per category and visualization layout. All plots in this chapter were made using the Matplotlib¹ library for Python. As it can be easily seen, the quantity of the insights was much higher for the matrix-based visualization layout. This is especially the case for categories 2, 4 and 7, which have a rather low domain value and thus represent simple insights. It can be said that the matrix-based visualization seems to be more suitable for simple insights, which are based on the study of individual nodes as well as links and related to attribute-based tasks. We think that this is due to the fact that the cell-based layout is particularly easy to handle and users are more invited to take a closer look at links with high weights, as they appear prominently as black boxes in the visualization. In the node-link diagram, links are represented as comparatively thin lines, furthermore they also often overlap

¹<https://matplotlib.org/stable/index.html>. Accessed: 2022-11-27.

with other links or nodes. As a result, users are less tempted to hover the mouse over these links to gain according insights about the exhibition locations. It was surprising for us that the same can be said about the nodes as well, since nodes are displayed quite prominently in node-link diagrams.

Category 3 insights relate to the dynamic network and refer to statements that outline the development of the network over the entire period. Here, our matrix-based visualization method using split matrices proved to be very useful. We were able to record considerably more insights in this category using the matrix-based layout.

While the difference between the two visualization layouts is marginal for category 5, the node-link layout turned out to be better for complex topology-related insights (category 8). These insights are very valuable in the art history domain according to the findings of our literature review.

7.3.3 Correct Insights Using the Map Visualizations

Insight categories 10 and 11 refer to the map visualization, which was the same for all participants. Due to our visualization design, many insights could only be gained with the help of the map. The map also proved to be a useful visualization layout to display the exhibitions, as the number of recorded insights is quite high.

The correct insights of categories 10 and 11 can be seen in Figure 7.3. Irrespective of the network visualization layout, almost exactly the same number of insights was obtained for the map view. Insights of category 10 are trivial and refer to insights that are immediately recognizable by looking at the map, such as statistics on individual exhibitions and statements about individual cities. More complex statements, which take the entire time period into account, are summarized in category 11.

Complex, comparative statements about the exhibition locations are summarized in category 5. In fact, all these statements were not made with the help of the map, but with the help of the network visualizations, which is why map-related insights can be summarized using only two categories.

7.3.4 Hypotheses

Two Participants were able to generate three hypotheses in total. As mentioned previously, we define hypotheses as insights that were made with the help of domain knowledge and go beyond the study of the visualizations. Two of these hypotheses are relatively simple and were discovered with the help of the map. One hypothesis is more complex and was made by analyzing the topology of the network. All hypotheses were made by the participants who used the node-link diagram. We counted all hypotheses as correct, a verification would require a corresponding literature research and is out of scope for this evaluation.

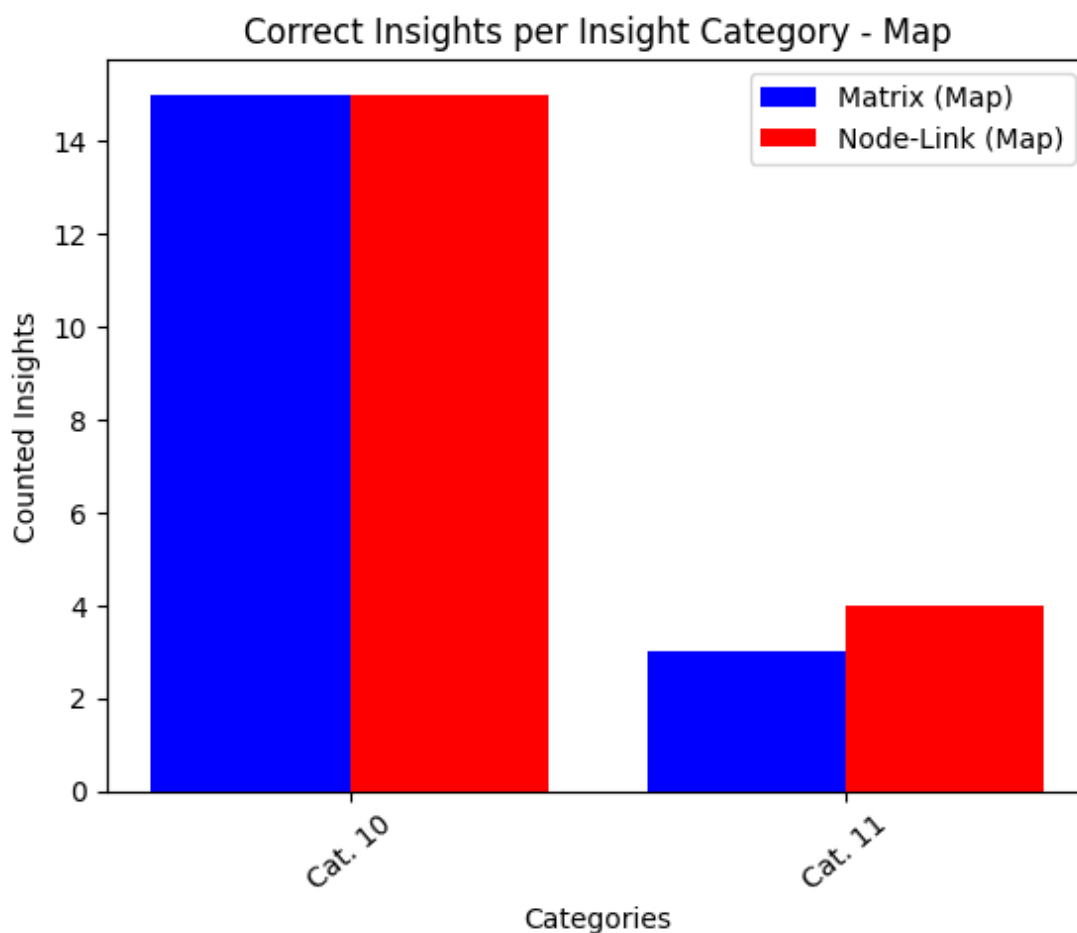


Figure 7.3: Correct insights gained using the map visualization.

7.3.5 Domain Value

Finally, we also quantitatively analyzed the domain values of the captured insights and hypotheses. To this end, we grouped insights according to their domain value. As before, we only included correct insights for our analysis. We considered all insights, including insights which were made using the map and insights which refer to hypotheses.

Statistical tests were performed in the previous adaptations of the insight-based evaluation method [LBS⁺19, SND04, GGZL14, NSD11]. Due to the small sample size and rare occurrences of some categories, we did not find it to be possible to perform meaningful statistical tests in most cases. We did use a *Mann-Whitney U test* [MW47] however to test whether the domain value of one visualization method tended to be higher than the other one. The Mann-Whitney U test is a non-parametric alternative to the t-test and is recommended when the (unpaired) data are not normal distributed [Har01, NH11].

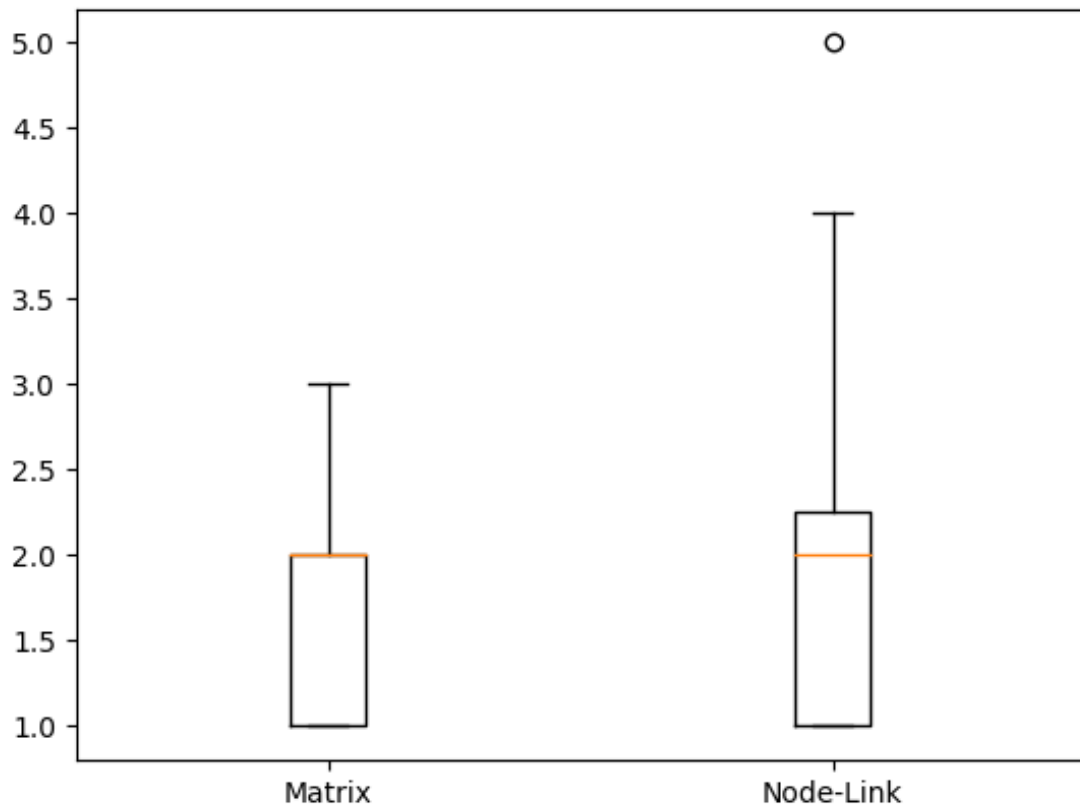


Figure 7.4: Boxplots of the domain value distributions.

Furthermore, the test can also be applied for ordinal data [Ran21], which applies to the Likert scale we use [SA13].

We performed the test using the SciPy² implementation. We obtained a rather high p-value of 0.31 (rounded to two decimal places), meaning that the observed difference is not statistically significant.

Two boxplots of the domain values can be seen in Figure 7.4. It is easy to see the outliers in the node-link diagram, which can be explained by the presence of the hypotheses. Mean domain values were 1.77 for the matrix-based visualization layout and 1.94 for the node-link diagram, respectively (rounded to two decimal places).

Based on these results, therefore, we cannot point to any notable differences between the two representations with respect to domain values.

²<https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.mannwhitneyu.html>. Accessed: 2022-11-27.

7.3.6 Comments during the Evaluation

We recorded a total of 14 comments on the application during the evaluation. The transcription of these can be found in the appendix. As all interviews had been conducted in German, the answers were translated afterwards.

Participants 1 and 2 were conducting the evaluation using the node-link diagram, participants 3 and 4 were using the adjacency matrix.

Node-Link Diagrams

We recorded three positive comments on the layout. Participant 1 highlighted that the application can be very valuable when researching an artist. Participant 2 found the study of temporal changes in the network to be particularly interesting. For example, it is possible to track how artists join groups and leave over time.

During the annual analysis of the network, participant 2 complained about the fact that the nodes are sometimes located in different places within the node-link diagram for different years. Thus, this is a mental map preservation problem [CP96, AP13], which we already discussed in detail in Chapter 3. The force-directed layout is responsible for the node movement throughout the years, because the coordinates of the nodes are recalculated each time based on the existing connections.

Participant 2 also noted that the importance of exhibitions cannot be determined simply on the basis of exhibition size. Rather, the importance of the works must also be considered.

Matrix-based Visualization

For the matrix-based evaluation, we recorded five positive comments. These were related to various aspects of the visualizations as well as the interaction functions. // // Participant 4 complained twice about the symmetry property of the adjacency matrix in the total view, as the double presence of entries caused a lot of confusion.

Map

Participant 4 generally liked the marker clustering function (as implemented in the leaflet library) in the map component.

Nevertheless, participant 4 was confused that both the individual locations (i.e., markers) and the clusters contained digits in the map component. The number in the clusters referred to the number of individual markers that were combined, while the number in the markers referred to the number of exhibitions for that particular city.

7.3.7 Interviews

We recorded a total of 12 minutes and 26 seconds for the interviews. The transcriptions can be found in the appendix. As mentioned before, participants 1 and 2 were conducting the evaluation using the node-link diagram, while participants 3 and 4 were using the matrix-based representation.

Opinion on the Visualization Method

Participants were generally satisfied with both visualization methods.

Participant 1 said that the node-link diagram is not self-explanatory and that the training session before the evaluation was needed. However, participant 1 also assumed that the application could be operated without further explanation after a certain period of use. Participant 2 was also satisfied with the node-link diagram, and felt that it could even be used directly in exhibitions, for example as a screenshot next to a hall text. Participants 3 and 4 were also satisfied with the adjacency matrix, but they were both bothered by the symmetry property in the total view. They were irritated by the fact that the same information appeared twice under the diagonal. Participant 3 also referred to the annual view, where both halves of the matrix were spatially separated - the participant would like to have a similar separation for the total view as well.

Suggestions for Improvement and Complaints

Participant 1 wished to have links to the DoME website, so that further data about the exhibitions could be viewed. Actually, this feature is already implemented, however it was not used during the evaluation due to the lack of time. Participant 2 wanted to have a better integration of the map component with the network visualization. In particular, the participant wished for a pop-up that would lead directly to the map. As mentioned before, participant 3 wanted to have a clear separation between the two matrix halves. Participant 4 wanted to see the names of the artists in the map view - currently, we simply reuse the map component without displaying further information. Also, the participant found the centrality-highlighting function in need of improvement, especially, when the names of the artists are colored by nationality. In particular, there were difficulties to identify the “medium-important” artists when using the centrality highlighting.

Suggestions for Improvement and Complaints

Regarding further visualization methods, two participants were already satisfied with the ones used currently. Participant 2 thought that displaying general statistics about the exhibitions would be interesting, but also pointed out that these can be found on the DoME [DoMb] website of the University of Vienna anyway. Participant 4, who had a quick look at the node-link diagrams before and after the evaluation, said that it would be a good idea to complement the matrix-based visualization with a node-link diagram.

In particular, participant 4 commented that topological information is difficult to extract from the matrix (“Who is the one that connects Belgium with Germany, yes?”).



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Discussion

In this Chapter, the sub-questions of our main research question are answered with respect to the evaluation results described in the previous Chapter 7. Furthermore, we also discuss usability-related issues of *Exhibitions Explorer*, which were identified in the evaluation, in a dedicated Section. Finally, we summarize the challenges regarding the insight-based evaluation protocol, as these can be of importance for future adaptations in the humanities domain.

8.1 Revisiting the Sub-Questions of this Thesis

Which state-of-the-art network visualization techniques or combination of such techniques are effective to help answering typical domain-relevant questions in art history research?

The first sub-question was mostly answered in the course of our SLR in Chapter 3.4. We comprehensively compared the state-of-the-art network visualization techniques with the current state of the art in the art history domain. It was shown that the study of basic [SRC17, LV20, McC19, Bur17, Kes63, NVCD22, Ric17] and more complex [LABV18, Swe13] node-link diagrams helped to answer domain-relevant questions. Nevertheless, there is a very large variety of visualization layouts and techniques which have not yet been adapted in digital art history.

Our application, *Exhibitions Explorer*, picks up on several methods we identified in the SLR. With the help of the design triangle framework by Miksch and Aigner [MA14], we were able to create a concept that can also be adapted easily in other applications in the art history domain.

In our view, the insight-based evaluation methodology proposed by Saraiya et al. [SND05] is best suited for proving the usefulness of visualization layouts and methods empirically.

By categorizing the insights, we were able to show *which* domain-relevant questions can be answered using a specific visualization design, and by analyzing the quantity and correctness of insights in the respective categories we could show *how well* these questions can be answered using a particular visualization.

To what extent can domain-relevant insights be gained through interactive visual data exploration?

In Chapter 2.2, we discussed how *Gephi* and *Palladio* in particular were successfully used in the art history domain to gain insights and generate hypotheses based on the visual exploration of networked data [SRC17, LV20, McC19, Bur17, Ric17, Con21].

However, accurately measuring such insights or hypotheses in the context of an evaluation represents a research gap in the art history domain. We therefore decided to adapt the insight-based evaluation approach to show that network visualizations, combined with suitable interaction functions, can lead to a large quantity and variety of different insights. Furthermore, we were also able to provide statements about the quality of these insights by using the well-known Likert scale.

We were able to note three hypotheses with very high domain values 4-5 as well as countless insights in varying complexity with domain values 1-3. Considering the more trivial insights, it should be said that visualizations promote them in a very efficient manner, which also led to the high occurrences in our evaluation. Moreover, simple insights mostly serve as foundations for more complex insights or hypotheses.

The categorization helped us to identify nine different insight categories, which could be obtained with the help of the network visualizations. Furthermore, we defined two additional insight categories related to the map view of the exhibitions. The participants also formulated three hypotheses with unique character.

This shows well that network visualizations - provided a suitable visualization design - allow much more than the study of topological structures. Multivariate networks can encode further information in nodes and links, which ultimately lead to more sophisticated insights. The visualization of dynamic networks allows art historians to make statements about the development of artists' networks over time.

Finally, it should be said that the art historians we cited in the SLR were able to examine their networks for a very long time. Furthermore, they were familiar with basic principles of network analysis and could therefore correctly classify the insights they made. In our very short evaluation time, we had to compromise and use a smaller network for evaluation purposes. Furthermore, it has to be mentioned that *precise* hypotheses in the art history domain must be made on the basis of literature, and visualizations can only provide support in this regard. Despite all these limitations, we were able to show that *Exhibitions Explorer* with its visualization design can be used to gain valuable, domain-relevant insights. All three visualization layouts we used were well-suited for this purpose.

Are matrix-based visualization layouts a suitable alternative to node-link layouts for studying small historical networks?

In task-based evaluation methods, the good performance of matrix-based representations compared to node-link diagrams was already shown for various tasks [GFC05, ABHR⁺13]. With regard to the art history domain, however, a more complex picture emerges.

Despite our small sample size, it became apparent that matrix-based visualizations were more inviting to explore the network at a detailed level, which ultimately led to a higher quantity of insights. In particular, links are prominently displayed, making them easier to examine. In case of multivariate networks, much information may be encoded in these links. Many created networks in the literature we found are based on network projections [Ric17, SRC17, LV20, Bur17]. In the course of network projection, it is possible to store comprehensive additional information within the created links in the form of properties. In this respect, we see a crucial advantage of using matrix-based representations: If a lot of domain-relevant information is stored as link properties, a matrix-based layout will facilitate the analysis of this information due to the prominent representation of links. Moreover, a surprising result emerged when we looked at how participants were studying the nodes of the network: Participants reported simple insights referring to node properties more often using the matrix-based representation. Again, it should be noted that potentially endless amount of information can be stored within these nodes. Thus, the simple access on a single node within the visualization facilitates the analysis of corresponding node properties.

Node-link diagrams, on the other hand, promote other insights better, especially of topological nature. In our evaluation, five distinctive complex claims were made about the topological structure of the evaluation network using the node-link diagram, while only one claim was made using the matrix-based visualization. As we found in our literature review, the study of such information is enormously helpful for art historians. Specifically, the betweenness centrality of nodes was often studied in the publications we found [SRC17, LV20, Ric17, McC19, Bur17]. Betweenness centrality can be quickly detected in smaller node-link diagrams *without* any highlighting, as these are mostly found in bridge structures. One participant correctly identified the French artist Girieud as the connecting link between the German and the Belgium sub-networks. The study of group structures is also facilitated in node-link diagrams by implicit grouping thanks to the force-directed layout. The study of group structures played an important role in the publications we studied [Ric17, McC19, Bur17, NVCD22]. To conclude, it is important to emphasize that the study of topological structures is very valuable in art history, which ultimately affects the choice of the visualization layout.

In our evaluation, we worked with a sparse network due to the limited training and evaluation time. Ghoniem et al. [GFC05] highlighted that matrix-based representations outperform node-link diagrams for dense networks in many tasks. To this end, it would be valuable to conduct the insight-based evaluation using denser networks in the future. Saraiya et al. [SNLD06] described a longitudinal study using the insight-based

methodology in an article. This protocol can also be adapted to evaluate our application. We leave this task for future work.

For dynamic networks, the mental map can be maintained extremely easily with matrix-based layouts, as the nodes stay fixed all the time. One of our participants, who used the node-link diagram, complained about the fact that for each year the nodes are always found in different places. However, this is difficult to avoid since attention is also to be paid to a clear arrangement of the nodes in the surface. With our newly proposed matrix-based visualization layout, we can maximize the efficient use of space for undirected networks and avoid mental map-related issues, as the artist names stay fixed in the surface. However, despite all the benefits of the matrix-based layout for dynamic networks, this does not change the previously discussed limitations for the study of more complex topological structures.

Finally, it should be stated that strengths and weaknesses of both visualization layouts can always be addressed by creating additional interaction as well as filtering options.

To make node-link diagrams useful for dense networks, applications can incorporate filters to reduce the density. Here, we found geographic as well as yearly filters to be helpful. They were also well-received by our participants. Again, we would like to emphasize that the choice of a suitable data model greatly simplifies the implementation of such filtering methods, which is why this should definitely be taken into account when developing visual analytics solutions for network analysis.

In case of matrix-based representations, the centralities can be highlighted by implementing appropriate algorithms. Also, arranging the cells based on group structures, which are determined algorithmically, can facilitate the analysis of topological structures. An example implementation is described by Mesquita [Mes18] in a blog article.

Therefore, if only one of the two visualization layouts is to be used, the decision strongly depends on the additional implementation effort one is willing to take on for further functionalities. Both layouts are easy to adapt due to numerous open source implementations on the web, but a high implementation effort must be considered if numerous complex functionalities are to be included.

In essence, we believe that both approaches complement each other well. We therefore encourage software developers to integrate *both* layouts in visual analytics solutions, especially in the humanities domain, to support end users in the best possible way for their research.

8.2 Usability Issues in Exhibitions Explorer

The insight-based evaluation approach proved to be enormously helpful for identifying usability issues. In particular, the spontaneous comments of the participants, which we did not explicitly ask them to provide, turned out to be very useful. During the subsequent semi-structured interview, some aspects that had already been mentioned

8.3. Challenges Regarding the Evaluation Protocol and Recommendations for Future Adaptions

before were repeated by the participants, even though some points were further specified. We believe that a final interview may not be necessary if a think-aloud protocol is used for the evaluation.

Essentially, we were able to identify most of the issues in the map component. The map component should display more information about the currently studied artists and their exhibitions, furthermore a more intuitive linking between the dashboard component and the map component needs to be implemented.

Additionally, it is important to pay attention to a spatial or other separation of the two matrix halves for the total view of undirected networks, since the symmetry property caused a lot of problems for the participants.

Finally, it should be said that the participants also gave spontaneous praise during the evaluation and that the identified usability problems were fortunately not causing any major issues. In particular, it should be emphasized that over 94% of the insights were correct.

8.3 Challenges Regarding the Evaluation Protocol and Recommendations for Future Adaptions

Conducting the insight-based evaluation required a significant amount of effort. Besides the transcription as well as the coding of the insights, the preparation and the selection of a suitable evaluation network turned out to be very difficult. Furthermore, capturing the individual insights was also very challenging. Deciding whether more complex insights could be broken down into their constituent parts or counted as a single unit was difficult in specific cases. As mentioned in Chapter 7, we decided to use an approach that counts simple insights and the more complex insights that build on them individually. However, despite the greatest possible rigor as well as transparency, the subjective nature of this evaluation method must not be underestimated.

We think that simply performing and summarizing statistical tests for quantified metrics are not sufficient to take full advantage of this evaluation method. Rather, they may even skew the results, since the sample size is usually small in these protocols, mainly due to the fact that domain experts are needed, who often have little time to spare. Furthermore, quantification and coding are subjective and difficult to understand without access to the primary research data. Categorizing the insights and analyzing them qualitatively, as North et al. [NSD11] did, increases the comprehensibility of the results. This makes it possible to analyze strengths as well as weaknesses of certain visualization methods qualitatively in the domain context, which in our view is the fundamental advantage of this approach over task-based methods.



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Conclusion

Based on the sub-questions we discussed in the previous chapter, we are able to answer our main research question, “*What are appropriate ways to adopt visual analytics for exploration of networked data and gaining insights in the domain of art history research?*”. By conducting parts of an SLR, we discussed various network visualization techniques, which are appropriate for network analysis in the art history domain and can be adapted in visual analytics solutions.

Based on these findings, we used the design triangle framework proposed by Miksch and Aigner [MA14] to design a prototype for the exploration of a historical social network. We combined three visualization layouts with interaction and coloring techniques, so that art historians can study the network with respect to space and time.

By adapting the insight-based evaluation proposed by Saraiya et al. [SND05], we were able to define, measure, quantify, and categorize the insights art historians were able to gain using our prototype. The high quantity, quality and variety of the insights we captured indicates that our application design is suitable to support art historians in their research. The visualization design can also be adapted for other historical networks.

9.1 Future Work

While we were able to answer the research question, one research gap we identified still requires more intensive research. There is also room for improvement in our application to provide even better support in the art history domain.

9.1.1 Noise Reduction for Networks in the Humanities Domain

As noted by Lamqaddam et al. [LABV18], art historians are more interested in smaller data sets, with under 50 individuals. As we highlighted in our work, reducing noise, i.e., unnecessary data, is by no means a trivial matter. For example, to identify artists

with high betweenness centrality, it may be necessary to include lesser known artists in a network. The risk of biases also increases, if certain artists are excluded from a particular visualization. Porras [Por17] explicitly warned about biases in data sets for the art history domain. While our application's network creation tool provides a way for art historians to create a small detailed network on their own, further qualitative research is needed to examine different approaches for noise reduction. Such an investigation is not only useful for art history, but may also reveal relevant possibilities across the entire humanities domain.

9.1.2 Exhibitions Explorer

Exhibitions Explorer proved to be a useful application in our evaluation. Nevertheless, there is still room for improvement, especially in terms of usability. In particular, the symmetry property of the undirected network should be emphasized more clearly in the matrix-based representation. The map component also caused some issues. One participant complained that the link to the map was not user-friendly, while another highlighted that the map did not display the names of the currently studied artists. The fact that both individual markers and clusters had numbers inside them also caused confusion, because these numbers referred to different things. We also think that the map component can be expanded further. For example, it should be possible to filter exhibitions by an artist. More filters are also needed in the network creation component to be able to reduce the size of the ego networks accordingly.

Further, more comprehensive evaluation is needed to prove the usefulness of the application for larger and denser networks. One possibility, for example, would be to conduct a field study, a method that Carpendale [Car08] also mentioned in her review. However, we would like to continue with the insight-based methodology here as well. An insight-based longitudinal study has already been conducted and described by Saraiya et al. [SNLD06]. It would also be valuable to further increase the number of evaluation participants for the longitudinal study.

9.2 Final Remarks

Visual analytics and network visualizations are helpful tools to support art historians in their research. However, we also have to agree with the more critical voices, who emphasize that it is always important to carefully reflect on the underlying data set and the visualizations based on them. Only the correct use of technical tools can provide added value.

Furthermore, in accordance with the literature we discussed, we would like to emphasize that both the design and evaluation of visual analytics solutions need to take important details of the respective application domain into account. The failure to do so may have a negative impact on the willingness of researchers to use such solutions for their own work.

Both problems can be addressed by incorporating more qualitative methods in the design and evaluation of visual analytics tools and by stronger interdisciplinary collaboration. Thus, we would like to end this thesis with the appeal to always develop technical tools with human needs in mind.



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

List of Figures

2.1	Three basic network visualization layouts.	10
3.1	Rottenhamer worked in Augsburg and Venice. These nodes are associated with the hometowns of the patrons as well as unknown collectors (represented as “UNK”), who purchased Rottenhamer’s works. The visualization was made by McCabe with Palladio [McC19].	17
3.2	Close-up of Rottenhamer’s social network, focusing on his relationship with Hieronymus Ott [McC19].	18
3.3	Single-step and multi-step animations to highlight changes in dynamic networks. [RM14] © 2014 IEEE	21
3.4	Timeline-based visualization using an adjacency list [HBW14]. © 2014 IEEE	22
3.5	A circular layout to visualize a network of cities [Hen13]. Modified.	23
3.6	Integrated visualization approach of a mesh network in Rostock, Germany [HSS11]. © 2011 IEEE	25
3.7	<i>NodeTrix</i> , a hybrid visualization approach [HFM07]. © 2007 IEEE	26
3.8	When coloring the cells in the matrix-based representation, the colors of the links in the node-link diagram change accordingly [HF06]. © 2006 IEEE .	27
3.9	Attribute-Driven Edge Bundling applied on aircraft trails over France [PHT15]. © 2015 IEEE	28
5.1	Final labeled property graph database model used for <i>Exhibitions Explorer</i> .	40
5.2	Web application architecture used for <i>Exhibitions Explorer</i>	42
6.1	Total view in the map component, displaying all exhibitions from 1904 - 1916.	47
6.2	Yearly view in the map component. In this Figure, only exhibitions in 1913 are displayed.	47
6.3	Overview of the network creation component.	49
6.4	Menu of the network creation component.	49
6.5	Navigation to another ego network. The previous ego is colored green, whereas the new ego is colored yellow.	50
6.6	The network selection allows to create networks by exploring ego networks of selected artists.	51
6.7	The nodes can be colored by sex.	51
6.8	The nodes can be colored by (primary) nationality.	52
		93

6.9	The modal displays all selected nodes of the created network.	52
6.10	Overview of the dashboard component.	55
6.11	Matrix: Total network view, displaying the network in the entire period. .	55
6.12	Matrix: Yearly network view, showing networks of two consecutive years.	56
6.13	Change of the yearly network. It can be seen that the network of 1905 stays fixed in the visualization.	57
6.14	Node-Link: Total network view, displaying the network in the entire period.	58
6.15	Node-Link: Yearly network view, displaying the network in 1910.	58
6.16	Combination of the filter and node coloring function. It can be seen that the British artists Adeney and Atkinson shared an exhibition in the UK. . . .	59
6.17	Complete (sub-)networks may indicate a key common exhibition.	60
6.18	The map component can be (re-)used to display common exhibitions on the map.	61
6.19	Weighted degree centralities of two networks in 1904 and 1905.	61
7.1	The network used for the insight-based evaluation, focusing on expressionist painters.	68
7.2	Correct insights gained using the network visualizations.	75
7.3	Correct insights gained using the map visualization.	77
7.4	Boxplots of the domain value distributions.	78

List of Tables

4.1	Three key entities and their properties.	32
7.1	Brief historical background of the 12 artists displayed in the evaluation network.	66
7.2	Insight categories, which are based on the 204 recorded insights.	72
7.3	Incorrect Insights. Grouped by Participant ID, Visualization Method and Insight Category	74
1	Appendix: Counted Insights per Insight Category	110



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Bibliography

- [ABHR⁺13] Basak Alper, Benjamin Bach, Nathalie Henry Riche, Tobias Isenberg, and Jean-Daniel Fekete. Weighted graph comparison techniques for brain connectivity analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, page 483–492, New York, NY, USA, 2013. Association for Computing Machinery.
- [AES05] Robert Amar, James Eagan, and John Stasko. Low-level components of analytic activity in information visualization. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pages 111–117, 2005.
- [AH98] Keith Andrews and Helmut Heidegger. Information slices: Visualising and exploring large hierarchies using cascading, semi-circular discs. In *Late Breaking Hot Topic Paper, IEEE Symposium on Information Visualization (InfoVis'98)*, pages 9–12, 1998.
- [AMST11] Wolfgang Aigner, Silvia Miksch, Heidrun Schumann, and Christian Tominski. *Visualization of Time-Oriented Data*. Springer London, 2011.
- [AP13] Daniel Archambault and Helen C. Purchase. The “map” in the mental map: Experimental results in dynamic graph drawing. *International Journal of Human-Computer Studies*, 71(11):1044–1055, 2013.
- [AR10] Jeffrey K. Aronson and Manoj Ramachandran. The diagnosis of art: Ernst ludwig kirchner’s ‘nervous breakdown’. *J R Soc Med*, 103(3):112–113, March 2010.
- [Art] Arthist.net. <https://arthist.net/>. Accessed: 2022-04-26.
- [Aze15] Hina Azeem. The art of edvard munch: a window onto a mind. *BJPsych Advances*, 21(1):51–53, 2015.
- [BAA⁺19] Koenraad Brosens, Jan Aerts, Klara Alen, Rudy Jos Beerens, Bruno Cardoso, Inez De Prekel, Anna Ivanova, Houda Lamqaddam, Geert Molenberghs, Astrid Slegten, Fred Truyen, Katlijne Van der Stighelen, and Katrien Verbert. Slow digital art history in action: Project cornelia’s computational approach to seventeenth-century flemish creative communities. *Visual Resources*, 35(1-2):105–124, 2019.

- [Bar] Jesús Barrasa. Rdf triple stores vs. labeled property graphs: What's the difference? <https://neo4j.com/blog/rdf-triple-store-vs-labeled-property-graph-difference/>. Accessed: 2022-09-11.
- [Bar16] Albert-László Barabási. *Network Science*. Cambridge University Press, 2016.
- [BBDW17] Fabian Beck, Michael Burch, Stephan Diehl, and Daniel Weiskopf. A taxonomy and survey of dynamic graph visualization. *Computer Graphics Forum*, 36(1):133–159, 2017.
- [Bos20] Mike Bostock. Mobile patent suits. <https://observablehq.com/@d3/mobile-patent-suits>, 2020. Accessed: 2022-09-17.
- [Bru05] Marie-Cécile Bruwier. Anna boch. 1 octobre -17 décembre 2000. *Les cahiers de Mariemont*, 32(1):116–116, 2005. Included in a thematic issue : Céramique.
- [Buc00] Mary Bucholtz. The politics of transcription. *Journal of Pragmatics*, 32(10):1439–1465, 2000.
- [Bur17] Daniel Burckhardt. Comparing disciplinary patterns: Exploring the humanities through the lens of scholarly communication. *Digital Humanities Quarterly*, 11(2), 2017.
- [CA20] Nadia Chouchani and Mourad Abed. Online social network analysis: detection of communities of interest. *Journal of Intelligent Information Systems*, 54(1):5–21, February 2020.
- [Car] Anto carte (1886 - 1954) - antoine carte. <https://ysebaert-louisseize-arts.com/artists/anto-carte/>. Accessed: 2022-09-10.
- [Car08] Sheelagh Carpendale. *Evaluating Information Visualizations*, pages 19–45. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.
- [CLY17] Siming Chen, Lijing Lin, and Xiaoru Yuan. Social media visual analytics. *Computer Graphics Forum*, 36(3):563–587, 2017.
- [Col21] Tom Collings. Controller-service-repository. <https://tom-collings.medium.com/controller-service-repository-16e29a4684e5>, 2021. Accessed: 2022-09-09.
- [Con21] Melanie Conroy. Networks, maps, and time: Visualizing historical networks using palladio. *Digital Humanities Quarterly*, 15(1), 2021.

- [CP96] Michael K. Coleman and D. Stott Parker. Aesthetics-based graph layout for human consumption. *Software: Practice and Experience*, 26(12):1415–1438, 1996.
- [Cra20] Harrison Cramer. Adjacency matrix. <https://bl.ocks.org/harrisoncramer/32bbcd8c360e6d8aa0d5b7a50725fe73>, 2020. Accessed: 2022-09-17.
- [DeS] Biografien - gustave de smet. <https://www.kettererkunst.de/bio/gustave-de-smet-1877.php>. Accessed: 2022-09-10.
- [DoMa] About the project. <https://exhibitions.univie.ac.at/info/project>. Accessed: 2022-09-18.
- [DoMb] Dome. <https://exhibitions.univie.ac.at/>. Accessed: 2022-03-06.
- [DoMc] Theresia elisa van hall. <https://exhibitions.univie.ac.at/person/520>. Accessed: 2022-09-10.
- [DPS02] Josep Díaz, Jordi Petit, and Maria Serna. A survey of graph layout problems. *ACM Comput. Surv.*, 34(3):313–356, sep 2002.
- [DV19] Melissa DeJonckheere and Lisa M. Vaughn. Semistructured interviewing in primary care research: a balance of relationship and rigour. *Family Medicine and Community Health*, 7(2), 2019.
- [EK10] David Easley and Jon Kleinberg. *Networks in Their Surrounding Contexts*, page 77–106. Cambridge University Press, 2010.
- [Eve] Jonathan Evens. Servaes, albert - vm - jonathan evens. <https://www.artway.eu/content.php?id=2178&lang=en&action=show>. Accessed: 2022-09-09.
- [Fel81] Scott L. Feld. The focused organization of social ties. *American Journal of Sociology*, 86(5):1015–1035, 1981.
- [Fra98] Andrew U. Frank. Different types of “times” in gis. In Max J. Egenhofer and Reginald G. Golledge, editors, *Spatial and Temporal Reasoning in Geographic Information Systems*, pages 40–62. Oxford University Press, Oxford, 1998.
- [GFC05] Mohammad Ghoniem, Jean-Daniel Fekete, and Philippe Castagliola. On the readability of graphs using node-link and matrix-based representations: A controlled experiment and statistical analysis. *Information Visualization*, 4(2):114–135, 2005.

- [GGZL14] Steven R. Gomez, Hua Guo, Caroline Ziemkiewicz, and David H. Laidlaw. An insight- and task-based methodology for evaluating spatiotemporal visual analytics. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pages 63–72, 2014.
- [Gir] Pierre girieud (1875 - 1948). <https://www.galeriepentcheff.co.uk/en/peintre-pierre-girieud>. Accessed: 2022-09-10.
- [GJFM18] Shawn Gu, John Johnson, Fazle E. Faisal, and Tijana Milenković. From homogeneous to heterogeneous network alignment via colored graphlets. *Scientific Reports*, 8(1):12524, August 2018.
- [Góm19] Sergio Gómez. *Centrality in Networks: Finding the Most Important Nodes*, pages 401–433. Springer International Publishing, Cham, 2019.
- [Gra73] Mark S. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380, 1973.
- [Gra76] Elspeth Graham. What is a mental map? *Area*, 8(4):259–262, 1976.
- [Gun18] Joachim Gunkel. Eine emanzipierte künstlerin. *Pädiatrie*, 30(6):70–70, December 2018.
- [GWG17] Chris Grodoski, Libba Willcox, and Samantha Goss. Visualizing historic networks in studies in art education. *Studies in Art Education*, 58(1):51–68, 2017.
- [Gó] John Alexis Guerra Gómez. Force-directed graph with link highlighting. <https://observablehq.com/@john-guerra/force-directed-graph-with-link-highlighting>. Accessed: 2022-09-09.
- [Har01] Anna Hart. Mann-Whitney test is not just a test of medians: differences in spread can be important. *BMJ*, 323(7309):391–393, August 2001.
- [HB05] Jeffrey Heer and Danah Boyd. Vizster: visualizing online social networks. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pages 32–39, 2005.
- [HBW14] Marcel Hlawatsch, Michael Burch, and Daniel Weiskopf. Visual adjacency lists for dynamic graphs. *IEEE Transactions on Visualization and Computer Graphics*, 20(11):1590–1603, 2014.
- [Hen13] Stefan Hennemann. Information-rich visualisation of dense geographical networks. *Journal of Maps*, 9(1):68–75, 2013.
- [HET12] Christophe Hurter, Ozan Ersoy, and Alexandru Telea. Graph bundling by kernel density estimation. *Computer Graphics Forum*, 31(3pt1):865–874, 2012.

- [HF06] Nathalie Henry and Jean-daniel Fekete. Matrixexplorer: a dual-representation system to explore social networks. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):677–684, 2006.
- [HFM07] Nathalie Henry, Jean-Daniel Fekete, and Michael J. McGuffin. Nodetrix: a hybrid visualization of social networks. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1302–1309, 2007.
- [HHS20] Marius Hogräfer, Magnus Heitzler, and Hans-Jörg Schulz. The state of the art in map-like visualization. *Computer Graphics Forum*, 39(3):647–674, 2020.
- [Hog11] Bernie Hogan. Chapter 11 - visualizing and interpreting facebook networks. In Derek L. Hansen, Ben Shneiderman, and Marc A. Smith, editors, *Analyzing Social Media Networks with NodeXL*, pages 165–179. Morgan Kaufmann, Boston, 2011.
- [Hop09] Talke Klara Hoppmann. Examining the ‘point of frustration’. the think-aloud method applied to online search tasks. *Quality & Quantity*, 43(2):211–224, March 2009.
- [HSS11] Steffen Hadlak, Hans-Jorg Schulz, and Heidrun Schumann. In situ exploration of large dynamic networks. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2334–2343, 2011.
- [HSSH20] Derek L. Hansen, Ben Shneiderman, Marc A. Smith, and Itai Himelboim. Chapter 3 - social network analysis: Measuring, mapping, and modeling collections of connections. In Derek L. Hansen, Ben Shneiderman, Marc A. Smith, and Itai Himelboim, editors, *Analyzing Social Media Networks with NodeXL (Second Edition)*, pages 31–51. Morgan Kaufmann, second edition, 2020.
- [HTL17] George F. Hurlburt, George K. Thiruvathukal, and Maria R. Lee. The graph database: Jack of all trades or just not sql? *IT Professional*, 19(6):21–25, 2017.
- [Hum] Humanities + design a research lab at stanford university. <https://hdlab.stanford.edu/>. Accessed: 2022-04-24.
- [Kes63] M. M. Kessler. Bibliographic coupling between scientific papers. *American Documentation*, 14(1):10–25, 1963.
- [Kie17a] Miriam Kienle. Between nodes and edges: Possibilities and limits of network analysis in art history. *Artl@s Bulletin*, 6(3), 2017.
- [Kie17b] Miriam Kienle. Digital art history “beyond the digitized slide library”: An interview with johanna drucker and miriam posner. *Artl@s Bulletin*, 6(3), 2017.

- [KMSZ09] Daniel A. Keim, Florian Mansmann, Andreas Stoffel, and Hartmut Ziegler. *Visual Analytics*, pages 3341–3346. Springer US, Boston, MA, 2009.
- [KPW14] Andreas Kerren, Helen C. Purchase, and Matthew O. Ward. *Introduction to Multivariate Network Visualization*, pages 1–9. Springer International Publishing, Cham, 2014.
- [LABV18] Houda Lamqaddam, Jan Aerts, Koenraad Brosens, and Katrien Verbert. Nahr: A visual representation of social networks as support for art history research. In *Proceedings of the 10th Nordic Conference on Human-Computer Interaction*, NordiCHI '18, page 634–641, New York, NY, USA, 2018. Association for Computing Machinery.
- [LBS⁺19] Andres Ledesma, Niranjana Bidargaddi, Jörg Strobel, Geoffrey Schrader, Hannu Nieminen, Ilkka Korhonen, and Miikka Ermes. Health timeline: an insight-based study of a timeline visualization of clinical data. *BMC Medical Informatics and Decision Making*, 19(1):170, August 2019.
- [LHT17] Antoine Lhuillier, Christophe Hurter, and Alexandru Telea. State of the art in edge and trail bundling techniques. *Computer Graphics Forum*, 36(3):619–645, 2017.
- [LNV⁺15] Andres Ledesma, Hannu Nieminen, Päivi Valve, Miikka Ermes, Holly Jimison, and Misha Pavel. The shape of health: A comparison of five alternative ways of visualizing personal health and wellbeing. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 7638–7641, 2015.
- [LPP⁺06] Bongshin Lee, Catherine Plaisant, Cynthia Sims Parr, Jean-Daniel Fekete, and Nathalie Henry. Task taxonomy for graph visualization. In *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization*, BELIV '06, page 1–5, New York, NY, USA, 2006. Association for Computing Machinery.
- [LV20] Stephanie C. Leone and Paul A. Vierthaler. Innocent x pamphilj’s architectural network in rome. *Renaissance Quarterly*, 73(3):897–952, 2020.
- [MA14] Silvia Miksch and Wolfgang Aigner. A matter of time: Applying a data–users–tasks design triangle to visual analytics of time-oriented data. *Computers Graphics*, 38:286–290, 2014.
- [Mac86] Jock Mackinlay. Automating the design of graphical presentations of relational information. *ACM Trans. Graph.*, 5(2):110–141, apr 1986.
- [Mac21] Megan E. Macken. An overview of recent digital humanities initiatives in us art libraries. *Art Libraries Journal*, 46(2):51–56, 2021.

- [Man16] Lev Manovich. The science of culture? social computing, digital humanities and cultural analytics. *Journal of Cultural Analytics*, 1(1), 5 2016.
- [McC19] Sophia Quach McCabe. Intermediaries and the market: Hans rottenhammer’s use of networks in the copper painting market. *Arts*, 8(2), 2019.
- [McG95] Joseph E. McGrath. Methodology matters: Doing research in the behavioral and social sciences. In Ronald M. Baecker, Jonathan Grudin, William A.S. Buxton, and Saul Greenberg, editors, *Readings in Human-Computer Interaction*, Interactive Technologies, pages 152–169. Morgan Kaufmann, 1995.
- [McM21] Caitlin McMullin. Transcription and qualitative methods: Implications for third sector research. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, September 2021.
- [Mer00] Russell Merris. *Graph Theory*. John Wiley & Sons, Ltd, 2000.
- [Mes18] Déborah Mesquita. Building a co-occurrence matrix with d3 to analyze overlapping topics in dissertations. <https://towardsdatascience.com/building-a-co-occurrence-matrix-with-d3-to-analyze-overlapping-topics-in-dissertations-fb2ae9470dee>, 2018. Accessed: 2022-09-10.
- [MGM⁺19] Fintan McGee, Mohammad Ghoniem, Guy Melançon, Benoit Otjacques, and Bruno Pinaud. The state of the art in multilayer network visualization. *Computer Graphics Forum*, 38(6):125–149, 2019.
- [MML07] Christopher Mueller, Benjamin Martin, and Andrew Lumsdaine. A comparison of vertex ordering algorithms for large graph visualization. In *2007 6th International Asia-Pacific Symposium on Visualization*, pages 141–148, 2007.
- [MPQ00] Jeremy L. Mennis, Donna J. Peuquet, and Liujian Qian. A conceptual framework for incorporating cognitive principles into geographical database representation. *International Journal of Geographical Information Science*, 14(6):501–520, 2000.
- [MW47] H. B. Mann and D. R. Whitney. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, 18(1):50 – 60, 1947.
- [Neoa] Betweenness centrality. <https://neo4j.com/docs/graph-data-science/current/algorithms/betweenness-centrality/#:~:text=1.-,Introduction,of%20nodes%20in%20a%20graph>. Accessed: 2022-09-18.

- [Neob] How to save the result of a query (sub graph) in cypher. <https://community.neo4j.com/t5/neo4j-graph-platform/how-to-save-the-result-of-a-query-sub-graph-in-cypher/m-p/25782/highlight/true#M12278/>. Accessed: 2022-09-11.
- [Neoc] Introduction. <https://neo4j.com/docs/graph-data-science/current/introduction/>. Accessed: 2022-09-16.
- [Neod] Neo4j bloom. <https://neo4j.com/product/bloom/>. Accessed: 2022-09-18.
- [Neoe] Projecting graphs using cypher. <https://neo4j.com/docs/graph-data-science/current/graph-project-cypher/>. Accessed: 2022-09-18.
- [Neof] Tutorial: Import relational data into neo4j. <https://neo4j.com/developer/guide-importing-data-and-etl/>. Accessed: 2022-09-14.
- [NH11] Barun K. Nayak and Avijit Hazra. How to choose the right statistical test? *Indian J Ophthalmol*, 59(2):85–86, March 2011.
- [NMSL19] Carolina Nobre, Miriah Meyer, Marc Streit, and Alexander Lex. The state of the art in visualizing multivariate networks. *Computer Graphics Forum*, 38(3):807–832, 2019.
- [Nor06] Chris North. Toward measuring visualization insight. *IEEE Computer Graphics and Applications*, 26(3):6–9, 2006.
- [NSD11] Chris North, Purvi Saraiya, and Karen Duca. A comparison of benchmark task and insight evaluation methods for information visualization. *Information Visualization*, 10(3):162–181, 2011.
- [NVCD22] Laura Noble, Valentina Vavassori, Alan Crookham, and Stuart Dunn. Networking the archive: The stories and structures of thos. agnew’s stock books. *J. Comput. Cult. Herit.*, 15(1), jan 2022.
- [Oko15] Chitu Okoli. A guide to conducting a standalone systematic literature review. *Communications of the Association for Information Systems*, 37(43), 2015.
- [Pae17] Arsenio Paez. Gray literature: An important resource in systematic reviews. *Journal of Evidence-Based Medicine*, 10(3):233–240, 2017.
- [Par19] Dmitry Paranyushkin. Infranodus: Generating insight using text network analysis. In *The World Wide Web Conference, WWW ’19*, page 3584–3589, New York, NY, USA, 2019. Association for Computing Machinery.

- [PHT15] Vsevolod Peysakhovich, Christophe Hurter, and Alexandru Telea. Attribute-driven edge bundling for general graphs with applications in trail analysis. In *2015 IEEE Pacific Visualization Symposium (PacificVis)*, pages 39–46, 2015.
- [Por17] Stephanie Porras. Keeping our eyes open: Visualizing networks and art history. *Artl@s Bulletin*, 6(3), 2017.
- [PVC21] Sumit Purohit, Nhuy Van, and George Chin. Semantic property graph for scalable knowledge graph analytics. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 2672–2677, 2021.
- [PW10] Victor R. Preedy and Ronald R. Watson, editors. *5-Point Likert Scale*, pages 4288–4288. Springer New York, New York, NY, 2010.
- [Qva16] Ludwig Qvarnström. The jewish modernist: Isaac grünewald in bertel hintze’s art history. *Journal of Art Historiography*, 15(LQ1), 2016.
- [Ran21] Priya Ranganathan. An introduction to statistics: Choosing the correct statistical test. *Indian J Crit Care Med*, 25(Suppl 2):S184–S186, May 2021.
- [RDP⁺13] Carolina Riveros, Agnes Dechartres, Elodie Perrodeau, Romana Haneef, Isabelle Boutron, and Philippe Ravaud. Timing and completeness of trial results posted at ClinicalTrials.gov and published in journals. *PLoS Med.*, 10(12):e1001566; discussion e1001566, December 2013.
- [Ric17] Yael Rice. Workshop as network: A case study from mughal south asia. *Artl@s Bulletin*, 6(3), 2017.
- [Ris17] Peter Risdon. August macke (1887–1914) and franz marc (1880–1916) in london. *The British Art Journal*, 18(2):56–59, 2017.
- [RM14] Sébastien Rufiange and Guy Melançon. Animatrix: A matrix-based visualization of software evolution. In *2014 Second IEEE Working Conference on Software Visualization*, pages 137–146, 2014.
- [SA13] Gail M. Sullivan and Anthony R. Artino, Jr. Analyzing and interpreting data from likert-type scales. *Journal of Graduate Medical Education*, 5(4):541–542, December 2013.
- [SBMK14] Martin Steiger, Jürgen Bernard, Thorsten May, and Jörn Kohlhammer. A survey of direction-preserving layout strategies. In *Proceedings of the 30th Spring Conference on Computer Graphics, SCCG ’14*, page 21–28, New York, NY, USA, 2014. Association for Computing Machinery.
- [Sch66] Ernst Scheyer. German expressionist prints, drawings and watercolors: Die brücke. *Bulletin of the Detroit Institute of Arts*, 45(3/4):43–45, 1966.

- [SKS96] Avi Silberschatz, Henry F. Korth, and S. Sudarshan. Data models. *ACM Comput. Surv.*, 28(1):105–108, mar 1996.
- [Sma73] Henry Small. Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24(4):265–269, 1973.
- [SND04] Purvi Saraiya, Chris North, and Karen Duca. An evaluation of microarray visualization tools for biological insight. In *IEEE Symposium on Information Visualization*, pages 1–8, 2004.
- [SND05] Purvi Saraiya, Chris North, and Karen Duca. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 11:443–456, 2005.
- [SNLD06] Purvi Saraiya, Chris North, Vy Lam, and Karen A. Duca. An insight-based longitudinal study of visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 12(6):1511–1522, 2006.
- [SRC17] Léa Saint-Raymond and Antoine Courtin. Enriching and cutting: How to visualize networks thanks to linked open data platforms. *Artl@s Bulletin*, 6(3), 2017.
- [Ste46] Stanley S. Stevens. On the theory of scales of measurement. *Science*, 103(2684):677–680, 1946.
- [Sua] Alejandro Suarez. (d3v4) force layout w highlights. <https://bl.ocks.org/almsuarez/4333a12d2531d6c1f6f22b74f2c57102>. Accessed: 2022-09-09.
- [Sve10] Patrik Svensson. The landscape of digital humanities. *Digital Humanities Quarterly*, 4(1), 2010.
- [Swe13] Robert W. Sweeny. Complex digital visual systems. *Studies in Art Education*, 54(3):216–231, 2013.
- [SYPB21] Sarah Schöttler, Yalong Yang, Hanspeter Pfister, and Benjamin Bach. Visualizing and interacting with geospatial networks: A survey and design space. *Computer Graphics Forum*, 40(6):5–33, 2021.
- [Val22] Gabriel Valiente. Adjacency maps and efficient graph algorithms. *Algorithms*, 15(2), 2022.
- [VBW17] Corinna Vehlow, Fabian Beck, and Daniel Weiskopf. Visualizing group structures in graphs: A survey. *Computer Graphics Forum*, 36(6):201–225, 2017.

- [VGSVZ⁺19] Guillermo Vega-Gorgojo, Laura Slaughter, Bjørn Marius Von Zernichow, Nikolay Nikolov, and Dumitru Roman. Linked data exploration with rdf surveyor. *IEEE Access*, 7:172199–172213, 2019.
- [vW06] Jarke J. van Wijk. Views on visualization. *IEEE Transactions on Visualization and Computer Graphics*, 12(4):421–432, 2006.
- [vW13] Jarke J. van Wijk. Evaluation: A challenge for visual analytics. *Computer*, 46(7):56–60, 2013.
- [WT17] Chaoli Wang and Jun Tao. Graphs in scientific visualization: A survey. *Computer Graphics Forum*, 36(1):263–287, 2017.



Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar
The approved original version of this thesis is available in print at TU Wien Bibliothek.

Appendix

Insights

Counted Insights per Category

Table 1 shows the counted insights per insight category. *ID* refers to the identifier of the insight category, *Category* refers to the textual description of the insight category, *DV* refers to the domain value (1 - 5), *CIM* refers to the correct insights using the matrix-based visualization, *CINL* refers to the correct insights using the node-link diagram, *IIM* refers to the incorrect insights using the matrix-based visualization, *IINL* refers to the incorrect insights using the node-link diagram, *TCI* refers to the total correct insights and *III* refers to the total incorrect insights.

Comments during Evaluation

This Section contains the transcriptions of all spontaneous comments made by the participants during the evaluation, both in German (original) and English (translated). *IV* refers to the interviewer, *P1* refers to participant 1, *P2* refers to participant 2, *P3* refers to participant 3 and *P4* refers to participant 4.

Original

P1 [25:18]: Also inhaltlich finde ich es ur spannend, weil man sich so vorhanteln kann. Das ist so bisschen wie, wenn man sich in ein Thema neu einliest, und man kommt über die – eigentlich durch die Literaturangaben durch einen Hinweis zum nächsten.

P1 [25:55]: Und ich glaube, dass es extrem wertvoll ist, wenn man jetzt konkret an einen Künstler, eine Künstlerin forscht.

P2 [05:45]: Dass die Namen entsprechend springen. Also, dass sie nicht auf einer – Das hat mich nämlich vorher kurz etwas verwirrt, weil dadurch wandert das, weil wenn man einen sucht, dann - normalerweise erwartet man, dass man an einer Stelle sich – [...] – Deswegen habe ich ihn vorher auch herausgezogen, weil wenn ich den betrachten möchte, dann natürlich sollte er in irgendeiner Weise an einer Stelle bleiben.

Table 1: Appendix: Counted Insights per Insight Category

ID	Category	DV	CIM	CINL	IIM	IINL	TCI	TII
1	General claim about relationships of the network or the network itself	1	8	3	0	2	11	2
2	General claim about relationships of the network or the network itself with respect to the year and / or location	2	33	19	3	0	52	3
3	Complex, comparative claim about relationships of the network over the entire period	3	6	2	0	0	8	0
4	General claim about the artists of the network	1	17	6	0	0	23	0
5	Complex, comparative claim about the cities / countries of the network	3	5	4	0	0	9	0
6	Identify if two artists had one or more common exhibitions	1	3	2	1	0	5	1
7	Identify if two artists had one or more common exhibitions with respect to the year and / or location	2	26	6	1	0	32	1
8	Complex claim regarding the topological structure of the network	3	1	5	0	0	6	0
9	Claim about the degree centralities in the network	2	4	3	0	0	7	0
10	General claim about exhibitions or exhibiting cities (map)	1	15	15	1	1	30	2
11	Complex claim about the exhibitions over the entire period (map)	3	3	4	0	1	7	1
12	Hypothesis about female artists in the network	5	0	1	0	0	1	0
13	Hypothesis about the expressionist's fame outside of the most fundamental locations	4	0	1	0	0	1	0
14	Hypothesis about the impact of artists' associations on the network	4	0	2	0	0	2	0

P2 [17:30]: "Wo man dann eigentlich auch ablesen kann, wann ist da was gegründet worden, wann haben sie die ersten Ausstellungen gehabt. Wie sich das langsam entwickelt das Netzwerk. Da sind jetzt erst einmal zwei, dann kommen immer mehr dazu. Die Gruppe wird größer, oder die Gruppe verändert sich nicht mehr. Oder irgendwer fällt aus der Gruppe raus."

P2 [25:15]: Das hängt jetzt davon ab, wie wichtig die Werke sind. Das ist jetzt nicht, dass man sagt – es gibt schon Ausstellungen, die quasi nach der Kataloganzahl klein erscheinen, aber dann schon für die Kunstgeschichte schon wichtige Werke drinnen haben.

P3 [1:09]: Guter Zeilenabstand, so dass ich die mal gut erkennen und lesen kann.

P3 [4:18]: Also, sehr sympathisch finde ich einmal diese farbliche Unterscheidung, das hilft – dass nicht nur Ziffern, Buchstaben hat – also die Farbe hilft da generell.

P4 [5:25]: Das hat mir aber wahnsinnig gut gefallen, also das mit der Zeitleiste, finde ich halt echt, einen guten – ein gutes AHA-Erlebnis, kann man sagen.

P4 [8:25]: Das finde ich auch gut, dass ich hier in die Tiefe gehen kann, und mir das genauer anschauen kann.

P4 [8:44]: Woran ich mich hier gewöhnen muss, ist dass die Information doppelt ist durch die Spiegelung.

P4 [9:20]: Also das ist für mich trotzdem immer dieser Umdenkprozess, ja, dass – man sieht nicht zwei Ausstellungen, sondern Achtung, es ist nur eine Ausstellung!

P4 [13:13]: Und das muss ich auch sagen, das hat mich ganz kurz ein bisschen verwirrt, weil die "1" sehr sehr groß ist. [...] Und die "2" ist kleiner.

P4 [16:45]: Aber, wenn ich da draufklicke, sehe ich das schon. Das finde ich – seh ich schon in die Richtung, in die es aufgeht. [...] Mhm, das finde ich super.

P4 [20:36]: Das war ganz klar Deutschland, Belgien, das war sehr gut visualisiert. Also das hat man sehr schön herauslesen können.

Translated

P1 [25:18]: So in terms of content, I think it's very exciting, because you can kind of work through it like that. It's a little bit like when you're reading into a new topic, and you come across the - actually through the literature references from one reference to the next one.

P1 [25:55]: And I think it's extremely valuable when you're conducting research specifically on an artist.

P2 [05:45]: That the names jump accordingly. So, that they are not on a - that's what confused me a little bit before, because that's how it moves, because if you're looking for one, then - normally you expect to find yourself in one place - [...] -that's why I pulled it

out before, because if I want to look at it, then of course it should stay in one place in some way.

P2 [17:30]: Where you can actually read when something was founded, when they had the first exhibitions. How the network slowly evolves. There are now at first two, then more and more come to it. The group gets bigger, or the group doesn't change anymore. Or somebody leaves the group.

P2 [25:15]: Now that depends on how important the works are. It's not that you can say - there are indeed exhibitions that seem small according to the number of catalogues, but then indeed have important works for art history.

P3 [1:09]: Good line spacing, so I can see and read them well.

P3 [4:18]: So, I find this color differentiation very appealing, that helps - that not only numbers, letters have - so the color helps there in general.

P4 [5:25]: But I really liked that, that timeline, I think it's a good - a good AHA moment, you can say.

P4 [8:25]: I think that's also good, that I can go in-depth here, and take a closer look.

P4 [8:44]: What I have to get used to here is that the information is duplicated by the mirroring.

P4 [9:20]: So that's - still - always this rethinking process for me, yes, that - you don't see two exhibitions, but attention, it's just one exhibition!

P4 [13:13]: And I have to say that too, that confused me a little bit for a very short time, because the "1" is very very big. [...] And the "2" is smaller.

P4 [16:45]: But, when I click on it, I see that already. I find that - I can already see the direction in which it opens [...] Mhm, I think that's great.

P4 [20:36]: That was clearly Germany, Belgium, that was very well visualized. So you could extract that very nicely.

Post-evaluation Interviews

This Section contains the transcriptions of all four post-evaluation interviews conducted with the participants, both in German (original) and English (translated). *IV* refers to the interviewer, *P1* refers to participant 1, *P2* refers to participant 2, *P3* refers to participant 3 and *P4* refers to participant 4.

Interview P1

Original

IV [0:04]: Also, die erste Frage, die ich hätte, was ist halt Ihre Meinung nur mal zur Visualisierungsmethode, also zu diesen Knoten mit diesen Strichen. Ist das aus Ihrer

Sicht eine gute Visualisierungsmethode für so etwas, fanden Sie das intuitiv?

P1 [0:21]: Also ich hätte mich ohne Erklärung nicht ausgekannt, ich habe mir das vorher ganz kurz angeschaut, und ich habe mir gedacht, ich verstehe es nicht [laughs]. Für mich war das nicht selbsterklärend, sagen wir es mal so – muss es auch nicht sein, ja, aber ich hätte das ohne Erklärung nicht wirklich umsetzen können. Vielleicht, wenn man sich eine Zeit lang damit spielt -

IV [0:36]: Ja.

P1 [0:36]: Kommt man sicher darauf.

IV [0:40]: Mhm.

P1 [0:41]: Aber jetzt rein vom Hinschauen alleine, hat es mir zu wenig gesagt, ja.

IV [0:45]: Okay, verstehe. Die nächste – gibt es etwas, das Sie verändern oder vielleicht – in dieser Applikation – verbessern würden, gibt es irgendwas, das Sie vermisst haben, irgendwas, was man immer noch genauer hätte herausarbeiten können, irgendein extra Filter oder sonst irgendwas?

P1 [1:03]: Ja, was ich interessant gefunden hätte, wäre es, wenn es zum Beispiel einen Link geben würde bei den Ausstellungen. Dass man dann erfährt eben, wie lang hat die gedauert, ob die nur wenige Wochen war oder ob die über Monate gelaufen ist.

IV [1:16]: Ja.

P1 [1:17]: Hat es da viel Besucher gegeben, Interesse – das wäre natürlich neben der Anzahl der Objekte und der Künstler aus meiner Sicht auch spannend, ja. Und natürlich, wie gut sind sie angenommen worden.

IV [1:27]: Und das letzte – gibt es vielleicht noch so eine Visualisierungsmethode, die sie zusätzlich – weil wir haben ja wirklich nur mit den Karten gearbeitet und mit diesen [points at Node-Link Diagram] – und gibt es vielleicht eine, was weiß ich, irgendein Bar Chart oder diese Kuchendiagramme oder was Sie sich gewünscht hätten zu statistischen Sachen, oder denken Sie sich, nein, das hat mir schon alles gepasst, um einen Eindruck zu bekommen.

P1 [1:51]: Ja, für den Eindruck reicht das völlig aus, absolut, finde ich schon, ja. Nein, also Tortendiagramme, die kannst du so – nein.

IV [1:59]: Okay, perfekt, vielen Dank, das wäre es auch schon gewesen.

Translated

IV [0:04]: So, the first question I would have, what is just your opinion just on the visualization method, so these nodes with these strokes. From your point of view, is that a good visualization method for something like that, did you find that intuitive?

P1 [0:21]: Well, I wouldn't have known my way around without an explanation, I looked at it very briefly beforehand, and I thought to myself, I don't get it [laughs]. It wasn't

self-explanatory to me, let's put it that way - it doesn't have to be, yeah, but I couldn't have really used it without an explanation. Maybe if you play with it for a while -

IV [0:36]: Yes.

P1 [0:36]: You come up with it for sure.

IV [0:40]: Mhm.

P1 [0:41]: But now purely from looking at it alone, it didn't say enough to me, yeah.

IV [0:45]: Okay, I see. The next - is there anything that you would change or maybe - in this application - improve, is there anything that you missed, anything that could have been elaborated on more, any extra filter or anything?

P1 [1:03]: Yes, what I would have found interesting if there was, for example, a link to the exhibitions. That you could then find out how long it lasted, whether it was only a few weeks or whether it ran for months.

IV [1:16]: Yes.

P1 [1:17]: Were there a lot of visitors, interest - that would of course also be exciting from my point of view, in addition to the number of objects and artists, yes. And of course, how well have they been received.

IV [1:27]: And the last one - is there maybe some kind of visualization method that additionally - because we really just worked with the maps and with these [points at node-link diagram] - and is there maybe a, I don't know, some bar chart or these pie charts or anything you would have liked for statistical things, or do you think, no, this has all been sufficient for me to get an impression.

P1 [1:51]: Yes, for the impression that's quite enough, absolutely, I think so, yes. No, so pie charts, you can - no.

IV [1:59]: Okay, perfect, thank you very much, that would have been it.

Interview P2

Original

IV [0:04]: Die erste Frage, die ich Ihnen noch gerne stellen würde ist – genau, was ist Ihre Meinung zu dieser Visualisierungsmethode, also zu diesen Knoten und Kanten. Haben Sie – Fanden Sie das intuitiv oder haben Sie sich da relativ gut zurechtgefunden?

P2 [0:21]: Also prinzipiell ja. Ich könnte mir auch vorstellen, dass wir so etwas dann entsprechend auch – was weiß ich - in Erklärungstafeln in Ausstellungen oder so etwas verwenden können, also je nachdem wie ich es da herausbekomme.

IV [0:35]: Mhm.

P2 [0:36]: Also ich kann das als Screenshot jetzt in der einfachsten Version nehmen und auf irgendeinen Saaltext darauf geben. Ist das sicherlich entsprechend interessant, da

kann man also dann auch was direkt in Ausstellungen damit anfangen, dass man sagt, okay, wie läuft denn jetzt diese Ausstellung und können sagen, er hat mit dem so und so oft ausgestellt, und das sehe ich jetzt grafisch aufbereitet.

IV [1:03]: Okay, vielen Dank. Dann, die nächste: Gibt es was, das Sie hier ändern oder verbessern würden, in der Applikation, gab es irgendwas, Sie haben ja vorher gemeint, dass Sie irgendetwas vermissen mit den genauen Ausstellungen oder –

P2 [1:18]: Genau, also einerseits diese Verknüpfung zwischen Karte

IV [1:24]: Ja.

P2 [1:25]: und den vorliegenden Spinnentieren.

IV [1:30]: Ja, ja.

P2 [1:32]: Irgendwie, dass man – ich weiß nicht, ob es vielleicht etwas einfach gestaltet werden kann – also, wenn ich sage, ich hab den, und dann poppt mir irgendwo unten auf, wo ich das gleich – also, wo ich nicht mehr so direkt hineingehen muss, sondern ich sag, ich hab das Jahr 1911, ich suche mir den Otto Mueller und da sehe ich dann auf der Karte – also, sozusagen dass es – also das ist jetzt halt wirklich auch im Grundstadium –

IV [1:58]: Ja, genau richtig.

P2 [1:59]: Aber, dass man dann etwas, sozusagen anwenderfreundlich agiert.

IV [2:07]: Okay, ja, das verstehe ich vollkommen. Und zu guter Letzt, gibt es glauben Sie noch eine Visualisierungsmethode, die Sie vermissen, zum Beispiel irgendwie ein Kuchendiagramm oder irgendwelche weiteren Statistiken, die Sie hier gerne gehabt hätten, zusätzlich zu dieser Karte und diesem Spinnennetz quasi.

P2 [2:28]: Das ist jetzt, ja, die Frage, man könnte natürlich noch vielleicht dann sagen, okay, dass man direkt die Ausstellung ansehen kann und man sagen kann, okay, da sieht man jetzt einmal, er hat, was weiß ich, fünf Werke drinnen gehabt, der nächste hat 10 Werke drinnen gehabt, der übernächste 25.

IV [2:48]: Ja.

P2 [2:48]: Oder so irgendetwas. Das ist jetzt, ja -

IV [2:52]: Mhm.

P2 [2:52]: Es hängt jetzt davon ab, wirklich jetzt vom – wobei man jetzt da in diesem weiteren Schritt der Uni Wien, mit der Homepage, gesehen hat -

IV [3:00]: Genau, ja ja.

P2 [3:00]: Also das müsste jetzt eigentlich, wenn man dorthin geht -

IV [3:03]: Ja.

P2 [3:04]: - eh auch klar sein.

IV [3:05]: Okay. Gut, dann vielen Dank, das waren auch schon die ganzen Fragen dahingehend.

Translated

IV [0:04]: The first question I would still like to ask you is - exactly, what is your opinion about this visualization method, that is, these nodes and linjs. Did you - Did you find that intuitive or did you feel relatively comfortable with that?

P2 [0:21]: So, in principle, yes. I could also imagine that we can use something like that; then accordingly also - what do I know - in explanation boards in exhibitions or something like that, so depending on how I can extract it from there.

IV [0:35]: Mhm.

P2 [0:36]: So I can take that as a screenshot now in the simplest version and put it next to some hall text. This is certainly interesting, because you can then also use it directly in exhibitions, that you say, okay, how is this exhibition going now and you can say, he exhibited with one so and so often, and I see that now graphically prepared.

IV [1:03]: Okay, thank you very much. Then, the next one: Is there anything that you would change or improve here, in the application, was there anything - you meant before that you miss something with the exact exhibitions, right -

P2 [1:18]: Exactly, so first this link between map

IV [1:24]: Yes.

P2 [1:25]: and the “arachnids”.

IV [1:30]: Yes, yes.

P2 [1:32]: Somehow, that one - I don't know whether it can perhaps be designed somewhat more simple - so, if I say, I have this, and then it pops up for me somewhere below, where I can see the same - so, where I no longer have to go into it that directly, but I can say, I study the year 1911, I look for Otto Mueller and there I finally see on the map - so, so to speak that it - so that is now really also in the basic stage -

IV [1:58]: Yes, exactly right.

P2 [1:59]: But, that one then acts somewhat, so to speak, user-friendly.

IV [2:07]: Okay, yeah, I totally understand that. And last but not least, do you think there's any visualization method that you miss, for example, somehow a pie chart or any other statistics that you would have liked to have here, in addition to this map and this “spider web”, so to speak.

P2 [2:28]: That's - now - yes, the question, you could of course still maybe then say, okay, that you can look directly at the exhibition and you can say, okay, there you can see, he had, what do I know, five works inside, the next one had 10 works inside, the one after that 25...

IV [2:48]: Yes.

P2 [2:48]: Or something like that. That's now, yes -

IV [2:52]: Mhm.

P2 [2:52]: It depends now, really now on the - whereby one has seen there in this further step of the University of Vienna, with the homepage.

IV [3:00]: Exactly, yes yes.

P2 [3:00]: So that should now, actually, if you go there -

IV [3:03]: Yes.

P2 [3:04]: - be clear anyway.

IV [3:05]: Okay. Well, thank you very much, that was all the questions.

Interview P3

Original

IV [0:08]: Die erste Frage wäre, mit dieser Visualisierungsmethode, mit dieser Matrix, was ist Ihre Meinung dazu, also hier geht es wirklich mal nur um diese Darstellung.

P3 [0:24]: Also ja, es ist das vertraute Koordinatensystem

IV [0:26]: Mhm.

P3 [0:27]: Das man aus dem Mathematik-Unterricht kennt, mit - kann jeder damit umgehen eigentlich.

IV [0:30]: Mhm.

P3 [0:31]: Also, das macht es mal sehr einfach und sehr klar. Mich hat das einzige - Wo ich ein bisschen gestolpert bin, war die Tatsache, dass ich immer „ah, da muss ich schauen, da ist noch was“...

IV [0:39]: Ja.

P3 [0:40]: Dass ich auf diese Spiegelung vergessen habe, wobei, es gibt eine Möglichkeit das auseinanderzuziehen, das war bei den Jahren, war das ja so.

IV [0:45]: Genau, genau. Ja.

P3 [0:46]: Ja, das ist das eine, was jetzt ein bisschen - das andere war, wenn das ein Block ist.

IV [0:53]: Ja.

P3 [0:53]: Dann habe ich immer, „ah ja, da muss ich“, und das war am Anfang, weil ich es nicht gewohnt war -

IV [0:57]: Mhm.

P3 [0:58]: Habe ich dann immer gedacht, ich muss da auch noch schauen.

IV [1:00]: Mhm.

P3 [1:01]: Dabei war es eigentlich nur das Doppelte, was ja nicht relevant war in dem Fall.

IV [1:07]: Okay.

P3 [1:08]: Oder, ja, da bin ich ein bisschen darüber gestolpert, weil ich immer geglaubt hab, „da, da muss ich jetzt auch noch einmal hinklicken“, dabei ist es ja dasselbe.

IV [1:16]: Ja, ja.

P3 [1:17]: Was ich schon in der anderen Diagonalen hatte.

IV [1:19]: Okay.

P3 [1:20]: Also so gesehen war es für mich – [laughs] nicht immer diesen Spalt dazwischen hätte. Wobei ich sagen muss, dass ich sowieso nicht so computeraffin bin, ich bin ja den Umgang mit digitalen Medien nicht so bewandert wie die jüngere Generation ganz einfach. Aber selbst ich habe mich eigentlich leicht getan, denn das mit dem Aufblättern, oder „rauf und runter“ klicken, das kennt man ja von jeder Homepage oder..

IV [1:46]: Mhm.

P3 [1:47]: Also es ist sehr bedienerfreundlich. Also, du wolltest noch eine andere Frage stellen.

IV [1:51]: Okay, die andere Frage, genau, etwas was Sie ändern oder verbessern würden in dieser Applikation, irgendwas was Ihnen vielleicht nicht so gut gefallen hat oder was man vielleicht noch einbauen könnte oder so.

P3 [2:09]: Ja, also wie gesagt, ich weiß nicht, ob es jetzt einen Sinn machen würde, da einen Strich durchzumachen.

IV [2:14]: Ja, mhm.

P3 [2:15]: Oder es so zu trennen, wie es bei den Jahren getrennt ist, weil – ich mein, das ist natürlich ein schönes Muster, aber eigentlich muss ich es.

IV [2:24]: Ja, okay.

P3 [2:24]: ja auseinanderdenken. Ja, das ist vielleicht ein geschlossenes Feld, aber -

IV [2:32]: Ja, ja.

P3 [2:32]: Ja, das wäre für mich vielleicht eine Erleichterung gewesen, wenn ich da noch einen Strich gehabt hätte oder es irgendwie räumlich getrennt hätten.

IV [2:42]: Mhm, verstehe. Und das letzte – Gibt es vielleicht noch eine Visualisierungsmethode, also vielleicht irgendein Kuchendiagramm oder sonst irgendwas, was Sie gerne gehabt hätten – etwas, das vielleicht - was das vielleicht noch gut ergänzen könnte. Wir haben ja jetzt mit dieser Matrix und mit der Karte gearbeitet, gibt es vielleicht eine dritte Visualisierungsmethode, die Ihnen jetzt einfällt, spontan?

P3 [3:04]: Karten, das ist man auch gewohnt, das kennt man, also Kuchen, du meinst diese Prozente?

IV [3:13]: Genau, genau, richtig.

P3 [3:14]: Also das ist mir persönlich nicht so wirklich sympathisch. Das kennt man von diversen Umfragen und so von Statistiken, und ja sicher, man sieht da auch - das ist ein ganz ein großer Spalt, aber da genügt mir eigentlich diese Farben, die Intensität der Farben,

IV [3:30]: Mhm.

P3 [3:31]: und das genügt mir da eigentlich. Da wüsste ich jetzt nicht noch etwas Zusätzliches. Wenn man sich einmal an das System gewöhnt hat mit den Kästchen und der Unterscheidung der Farben und rechts dann nur die Zusatzinformation, also ich fand das ist so graphisch gut gelöst und da geht mir jetzt nicht ab, dass ich da noch was Zusätzliches nicht habe.

IV [3:55]: Mhm, okay, verstehe. Gut, vielen Dank, das war es eh auch schon.

Translated

IV [0:08]: The first question would be, with this visualization method, with this matrix, what is your opinion about it, so this is really just about this representation for now.

P3 [0:24]: So yes, it's the familiar coordinate system.

IV [0:26]: Mhm.

P3 [0:27]: That you know from mathematics lessons, with - everyone can actually deal with it.

IV [0:30]: Mhm.

P3 [0:31]: So, that makes it very simple and very clear. The only - where I struggled a little bit was the fact that I was always like "ah, I have to look, there's something else".

IV [0:39]: Yes.

P3 [0:40]: That I forgot about this reflection, whereas, there is a way to pull that apart, that was with the years, was that so.

IV [0:45]: Exactly, exactly. Yes.

P3 [0:46]: Yeah, that's the one thing now that was a little bit - the other thing was, if that's a block.

IV [0:53]: Yes.

P3 [0:53]: Then I was always like "ah yeah, I have to" and that was in the beginning, because I wasn't used to it.

IV [0:57]: Mhm.

P3 [0:58]: I always thought then, I have to look there too -

IV [1:00]: Mhm.

P3 [1:01]: - but it was actually only the double, which wasn't relevant in that case.

IV [1:07]: Okay.

P3 [1:08]: Or - yes, I struggled with it a bit, because I always thought, "there, there I have to click again," but it's the same thing.

IV [1:16]: Yes, yes.

P3 [1:17]: Which I already had in the other diagonal.

IV [1:19]: Okay.

P3 [1:20]: So, from that point of view, it was for me - [laughs] not always this gap in between. Whereby, I have to say that I'm not that computer-savvy anyway, I'm not as proficient in dealing with digital media as the younger generation, simply put. But even I had an easy time, because you know that with the page scrolling, or clicking "up and down", you know that from every homepage, right..

IV [1:46]: Mhm.

P3 [1:47]: So it's very user friendly. So, you wanted to ask another question.

IV [1:51]: Okay, the other question, exactly, anything that you would change or improve in this application, anything that maybe you didn't like so much or that maybe could be added or something.

P3 [2:09]: Yes, so as I said, I don't know, if it would make sense to draw a line through it now.

IV [2:14]: Yeah, mhm.

P3 [2:15]: Or to separate it the way it's separated with the years, because - I mean, that's a nice pattern, of course, but actually I have to -

IV [2:24]: Yeah, okay.

P3 [2:24]: Yeah, think apart. Yeah, that might be a closed field, but...

IV [2:32]: Yeah, yeah.

P3 [2:32]: Yeah, that might have been a relief for me if I had another line there or somehow separated it spatially.

IV [2:42]: Mhm, I see. And the last - Is there perhaps another visualization method, so perhaps some pie chart or something else that you would have liked to have - something that could perhaps complement this well. We have now worked with this matrix and with the map, is there perhaps a third visualization method that occurs to you now, spontaneously?

P3 [3:04]: Maps, you are also used to that, you know these, so pie, you mean these percentages?

IV [3:13]: Exactly, exactly, right.

P3 [3:14]: Well, I personally don't really like that. You know that from various surveys and from statistics, and yes sure, you can see that - that's quite a big gap, but actually these colors, the intensity of the colors, it's enough for me -

IV [3:30]: Mhm.

P3 [3:31]: - and that is actually enough for me. I wouldn't know of anything additional now. Once you get used to the system with the boxes and the differentiation of the colors and then only the additional information on the right, well, I thought that was graphically well solved and I don't see that - I don't have anything additional.

IV [3:55]: Mhm, okay, I see. Well, thank you very much, that was it anyway.

Interview P4

Original

IV [0:05]: Was ist jetzt Ihre allgemeine Meinung zu dieser Kästchen-Visualisierungsmethode, zu dieser Matrix?

P4 [0:11]: Ich finde es schön mit diesem Anklicken, ich find schön, wo sich die Farben herausheben - jetzt bei den unterschiedlichen Filtermöglichkeiten, das gefällt mir gut. Wo ich mich schwer tue, ist, dass es gespiegelt ist, das heißt, man nimmt am ersten Blick Sachen doppelt wahr, man glaubt das ist doppelt so oft, wie es tatsächlich ist. Also mit dem habe ich mich bisschen schwer getan.

IV [0:31]: Mhm.

P4 [0:34]: Was interessant ist, was ich gut finde, ist, wenn ich dann anfangs mit dem Filtern, dann kommen so klar eben die Städte, bisschen diese Schwerpunkte, diese Nationalitäten, aufgrund der unterschiedlichen Farben dann heraus. So kann man das eigentlich gut zuordnen. Das wäre so mein - also, das, was für mich am schwierigsten war, ist das Doppelte wegzunehmen.

IV [0:55]: Okay. Gut, die zweite Frage ist (...) Genau, die zweite Frage ist, gibt es etwas, was Sie gerne verändern oder verbessern würden an dieser existierenden Applikation jetzt?

P4 [1:13]: Ich weiß nicht, ob das geht, weil man ja bei der Mappe - bei der Map -

IV [1:19]: Ja.

P4 [1:23]: Bei der Landkarte, wo alles zu sehen ist, für mich, ich hätte gerne oben noch weiter stehen, welche 2 Künstler ich jetzt gerade vergleiche miteinander, zum Beispiel. Also, dass man da noch diesen einen Hinweis hat. Das wäre mir eingefallen. Und das

eine, was noch war, war dass diese zwei Grüntöne relativ nahe beieinander waren, das habe ich bei dem ersten Blick nicht erkannt, dass einer ja auch noch hervorgehoben wird.

IV [1:41]: Mhm.

P4 [1:41]: Mit dem Fett, also da hat sich die Schrift verändert glaube ich von normal auf fett, und das dunklere Grün hat man ja auch noch dran gehabt, so dass ich beim helleren Grün nicht gemerkt habe, dass es auch auf fett gegangen ist. Das wären so meine –

IV [2:00]: Okay, super. Und das letzte, was ich Ihnen noch gerne fragen würde, ist, genau, ob es eine spezielle Visualisierungsmethode gäbe, zum Beispiel ein Kuchendiagramm oder irgendwas, was Sie noch gerne hätten, was diese zwei, Matrix und Mappe, noch ergänzen könnte?

P4 [2:19]: Ja, also ich fände es schön, das Netzwerk wirklich als Netzwerk zu sehen, also diese Verbindungen. Da wäre es vielleicht für mich gar nicht in erster Linie so – ja, also ist auch schön, dass man sieht wer ist von der Häufigkeit - miteinander verglichen, aber, wirklich, dass man dann sieht – ich tue mir hier schwer, herauszufinden, wer ist jetzt der eine Knotenpunkt. Wer ist der, der Belgien mit Deutschland verbindet, ja? Und das wäre in so einem Netzwerk, als Netz, vielleicht auch gut, als Ergänzung. Ich finde, es ergänzen sich beide Teile gut.

IV [2:59]: Okay, alles klar. Super, vielen Dank.

Translated

IV [0:05]: Now, what is your general opinion about this box visualization method, about this matrix?

P4 [0:11]: I think it's nice with this clicking, I like where the colors stand out - now with the different filtering options, I like that. Where I have a hard time with, is that it's mirrored, that means at first glance you perceive things twice, you think that's twice as often as it actually is. So I struggled with that a little bit.

IV [0:31]: Mhm.

P4 [0:34]: What's interesting, what I consider good, is when I start with the filtering, then the cities come out clearly, these focal points, these nationalities, because of the different colors. So, you can actually put that in relation well. That would be my - well, the thing that was most difficult for me is to take away the double.

IV [0:55]: Okay. Alright, the second question is (...) Exactly, the second question is, is there anything that you would like to change or improve about this existing application now?

P4 [1:13]: I don't know if that's possible, because with the "Mappe" [German] - with the map - you know.

IV [1:19]: Yes.

P4 [1:23]: With the map, where everything can be seen, for me, I would like to have displayed at the top, which two artists am I comparing with each other now, for example. So, that one still has this one reference there. That would have occurred to me. And the one thing that was, that, these two green tones were relatively close to each other, I didn't recognize that at the first glance, that one is also still highlighted.

IV [1:41]: Mhm.

P4 [1:41]: With the bold, so there the font changed I think from normal to bold, and you still had the darker green, so I didn't notice with the lighter green that it also went bold. So these would be my -

IV [2:00]: Okay, great. And the last thing I'd like to ask you is, exactly, if there was a specific visualization method, for example, a pie chart or anything else that you'd like to have, that could complement these two, matrix and map?

P4 [2:19]: Yeah, so I would like to see the network really as a network, so these connections. There it would be - perhaps for me, primarily, not at all - yes, so it is also nice that one sees who is from a frequency point of view - compared with each other, but, really, that one sees then - I have a hard time here to find out, who is now the one node. Who is the one that connects Belgium with Germany, yes? And that would be in a network, as a network, maybe also good, as a complement. I think both parts complement each other well.

IV [2:59]: Okay, all right. Great, thank you very much.